REPORT RESUMES

ED 003 289   24
DEVELOPMENT OF A K SCALE FOR A MOTIVATIONAL TEST BATTERY.
BY- THWEATT, ROGER C.
WESTERN STATE COLL. OF COLORADO, GUNNISON
REPORT NUMBER CRP-1594   PUB DATE JUN 65
EDRS PRICE MF-$0.18 HC-$5.04   126P.

DESCRIPTORS- *MOTIVATION, STUDENT MOTIVATION, MOTIVATION TECHNIQUES, *ATTITUDES, EDUCATIONAL ATTITUDES, *TESTING, TEST CONSTRUCTION, EVALUATION, MEASUREMENT TECHNIQUES, PREDICTIVE VALIDITY, *UNDERACHIEVERS, *OVERACHIEVERS, STUDENTS, HIGH SCHOOL STUDENTS, GRADE 11, GUNNISON, COLORADO, MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI)

THE PROBLEM OF THIS RESEARCH WAS TO DEVELOP AND VALIDATE A K SCALE FOR A MOTIVATIONAL TEST BATTERY. APPROXIMATELY 4,200 11TH GRADE, MICHIGAN PUBLIC HIGH SCHOOL STUDENTS COMPRISED THE STUDY. PROTOCOLS OF ABOUT 300 IDENTIFIED OVER- AND UNDERACHIEVERS WERE USED. AN ANALYSIS WAS CONSTRUCTED TO IDENTIFY ITEMS ON WHICH UNDERACHIEVERS SCORE AS OVERACHIEVERS AND OVERACHIEVERS SCORE AS UNDERACHIEVERS. BOTH TYPES OF ACHIEVERS WERE RANDOMLY ASSIGNED BY SEX TO A VALIDATION AND A CROSS-VALIDATION SAMPLE. EVIDENCE OF VALIDITY AND RELIABILITY OF THE K SCALE WAS EXAMINED BY--(1) APPLYING THE K SCALE TO ALL MISCLASSIFIED OVER- AND UNDERACHIEVERS, (2) DETERMINING THE EFFECTS ON MOTIVATIONAL BATTERY DISCRIMINATING ITEM RELIABILITY BEFORE AND AFTER APPLICATION OF THE MALE AND FEMALE K SCALE, AND (3) DETERMINING THE EFFECTS ON MOTIVATIONAL BATTERY VALIDITY WITH STANDARDIZED GRADE-POINT AVERAGES BEFORE AND AFTER APPLICATION OF THE MALE AND FEMALE K SCALE. THE CONCEPT UNDERLYING THE K SCALE OF THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY WAS SUBSTANTIATED FROM THE STANDPOINT OF THE RATIONALE OF DEFENSIVE TEST-TAKING ATTITUDES. IT WAS CONCLUDED THAT THE K CONCEPT POSSESSES THE POTENTIAL FOR A DISCRIMINATION SCALE BETWEEN OTHER SUBTLE BEHAVIORAL PHENOMENA. FURTHER INVESTIGATION WAS RECOMMENDED, HOWEVER, BEFORE ACTUAL EMPLOYMENT IN TEST BATTERY INTERPRETATION. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Development of a K Scale for a Motivational Test Battery

Final Report
of
Cooperative Research Project Number 1594

Roger C. Thweatt
Project Director

Western State College
Gunnison, Colorado

June, 1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, United States Department of Health, Education and Welfare

ED 003289

Acknowledgements

Four people played a significantly important role in the initiation and completion of this project.

Deep appreciation is expressed to this investigator's chairmen and fellow colleagues in the academic pursuit during the duration of the investigation: Dr. Kenneth Hansen, Director of the School of Education, Western State College of Colorado who provided the necessary support and encouragement to motivate this writer's initiation of the project; and Dr. Charles Fauset, Dean of the College of Education, Northern Arizona University who provided the impetus and means to bring the project to completion through many setbacks and disappointments. Many thanks go out to these two dedicated and capable men.

Without the astute research design knowledge and penetrating statistical insights (which invade brilliancy) of William Farquhar, College of Education, Michigan State University this investigation could never have been brought to fruition. Sincerest thanks, admiration and appreciation are extended to Dr. Farquhar who is responsible for this investigator's journey into the academic world.

Finally, the wife of most any researcher will be extensively although indirectly involved with his frustrations and concerns in pursuing completion of a long-term project. Such was the case of this investigator's wife who was directly

involved in many ways and who never faltered in her support, encouragement, and cooperation in seeing the end of the project. Deepest and warmest admiration are extended to her.

RCT

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS.................................... ii

LIST OF TABLES...................................... vi

LIST OF FIGURES..................................... vii

CHAPTER

I. THE PROBLEM.. 1
   Statement of the Problem........................ 10
   Background of Theory and Research............... 10
   Summary Statement of Organization............... 24

II. PREVIOUS ATTEMPTS TO DEVELOP VALIDITY KEYS: A REVIEW OF THE LITERATURE.......................... 25
   The L Scale..................................... 25
   The F Scale..................................... 29
   The K Scale..................................... 36
   Subtle-Obvious Keys (S-O)....................... 54
   Set T Scale..................................... 59
   The B Scale..................................... 63
   Miscellaneous Scales............................ 65
   Similarity Between Various Scales............... 70
   Summary... 76

III. PROCEDURES....................................... 78
   Background of Farquhar's Study.................. 78
   General Design of the Motivational Study....... 79
   Instrumentation................................. 82
   Procedures for the Present Investigation....... 83

IV. RESULTS OF THE INVESTIGATION...................... 86
   Distribution of K Items......................... 86
   Reliability of the K Scale...................... 88
   Validation of the K Scale....................... 89
   Effect of K on Validity......................... 91
   Summary... 92

V. SUMMARY AND CONCLUSIONS............................ 96
   The Problem..................................... 96
   Methodology..................................... 96
   Conclusions..................................... 99
   Implications for Further Research.............. 101

PAGE

BIBLIOGRAPHY 102

APPENDIX A.. 112

APPENDIX B.. 115

LIST OF TABLES

TABLE PAGE

I. Intercorrelations of K with Other MMPI Variables.. 48

II. Correlations of K Scale with Other Variables Thought to Be Loaded with the K-Factor............ 50

III. Intercorrelations of Five Scales Thought to Be Loaded with the Test-Taking Attitude, No Item Overlap, n=150 Normal Males........................ 50

IV. Intercorrelations of Four MMPI Validity Scale Indicators and Hy for Normals...................... 71

V. Summary of Theory of Need-Achievement and Non-Need Achievement Motivation Basic to Current Research.. 78

VI. Hypothesized Personality Factors Associated with Academic Achievement............................. 79

VII. Number of K Items in Each Test with Sex Differences in Selection........................... 86

VIII. Effects on M-Scale Internal Consistency Reliability Before and After Application of the Male and Female K Scale for the Validation Group........... 90

IX. Effects on M-Scales Internal Consistency Reliability Before and After Application of the Male and Female K Scale for the Cross-Validation Group.. 90

X. Effects on the Validity Coefficient Between M-Scales Raw Scores and Standardized Grade Point Averages After Application of the Male and Female K Scale for the Validation and Cross-Validation Groups......................... 92

LIST OF FIGURES

FIGURE PAGE

I. Methodological Selection of Individuals with Stable Measured Aptitude.......................... 81

II. Method of Selecting Under- and Over-Achievers.... 81

III. Theorized Model of Selection of Misclassified Under-Achievers and Over-Achievers Differentiated by the M-Scales.................................... 83

IV. Validation and Cross-Validation K Item Frequency Distribution for Males......................... 87

V. Validation and Cross-Validation K Item Frequency Distribution for Females............... 88

CHAPTER I

THE PROBLEM

The most important failing of almost all structured objective tests is their susceptibility to faking or lying. In addition, objective tests possess an even greater susceptibility to unconscious self-deception and role playing.[1] The possibility of such factors having an invalidating effect upon scores has been noted by many writers.[2-26] One of the assumed advantages of the projective methods is that they are relatively less

---

[1] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:525-564.

[2] C. R. Adams, "A New Measure of Personality," Journal of Applied Psychology, 1941, 25:141-151.

[3] G. W. Allport, "A Test for Ascendance-Submission," Journal of Abnormal Psychology, 1928, 23:118-136.

[4] G. W. Allport, "The Use of Personal Documents in Psychological Science," Social Science Research Council Bulletin, 1942, Number 42.

[5] R. G. Bernreuter, "Validity of the Personality Inventory," Personality Journal, 1933, 11:383-386.

[6] R. G. Bernreuter, "Theory and Construction of the Personality Inventory," Journal of Social Psychology, 1933, 4:387-405.

[7] R. G. Bernreuter, "The Present Status of Personality Trait Tests," Educational Research Supplement, 1940, 21:160-171.

[8] Marion Bills, "Selection of Casualty and Life Insurance Agents," Journal of Applied Psychology, 1941, 25:6-10.

[9] E. S. Bordin, "A Theory of Vocational Interests as Dynamic Phenomena," *Educational and Psychological Measurement*, 1943, 3:49-65.

[10] P. Eisenberg and A. Wesman, "A Consistency in Responses and Logical Interpretation of Psychoneurotic Inventory Items," *Journal of Education Psychology*, 1941, 32:321-338.

[11] J. P. Guilford and R. B. Guilford, "Personality Factors S, E, and M and Their Measurement," *Journal of Psychology*, 1936, 2:109-127.

[12] D. G. Humm and K. A. Humm, "Validity of the Humm-Wadsworth Temperament Scale: With Consideration of the Effects of Subjects' Response-Bias," *Journal of Psychology*, 1944, 18:55-64.

[13] D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Scale," *American Journal of Psychiatry*, 1935, 91:163-200.

[14] E. L. Kelley, C. C. Miles and L. M. Terman, "Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality," *Character and Personality*, 1936, 4:206-215.

[15] D. A. Laird, "Detecting Abnormal Behavior," *Journal of Abnormal Psychology*, 1926, 20:128-141.

[16] C. Landis and S. E. Katz, "The Validity of Certain Questions Which Purport to Measure Neurotic Tendencies," *Journal of Applied Psychology*, 1934, 18:343-356.

[17] J. B. Maller, "The Effect of Signing One's Name," *School and Society*, 1930, 31:882-884.

[18] W. C. Olson, "The Waiver of Signature in Personal Reports," *Journal of Applied Psychology*, 1936, 20:442-450.

[19] Saul Rosenzweig, "A Suggestion for Making Verbal Personality Tests More Valid," *Psychological Review*, 1934, 41:400-401.

[20] Saul Rosenzweig, "A Basis for the Improvement of Personality Tests with Special Reference to the M-F Battery," *Journal of Abnormal and Social Psychology*, 1938, 33:476-488.

[21] F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type Personality Test," In Quinn

influenced by such distorting factors. However, even on projectives malingerers perform in a discriminable fashion. (Here the clues are extreme cautiousness and hesitancy, rejection of cards, and a minimization of response in general.)[27]

The existence of a distorting influence in test-taking attitude is so obvious that it has hardly been thought necessary to establish it experimentally. However, a number of investigations have empirically demonstrated the effect. Frenkel-Brunswik investigated tendencies to self-deception in rating oneself, finding in some cases marked negative relations between self-judgments and the evaluation of others.[28] Hendrickson reported

---

McNemar and M. A. Merrill, Studies in Personality (New York: McGraw-Hill, 1942), pp. 229-234.

[22] E. K. Strong, Vocational Interests of Men and Women (Stanford: Stanford University Press, 1943).

[23] P. M. Symonds, Diagnosing Personality and Conduct (New York: Appleton-Century, 1932).

[24] P. E. Vernon, "The Attitude of the Subject in Personality Testing." Journal of Applied Psychology, 1934, 18:165-177.

[25] J. N. Washburne, "A Test of Social Adjustment." Journal of Applied Psychology, 1935, 19:125-244.

[26] R. R. Willoughby and M. E. Morse, "Spontaneous Reactions to a Personality Inventory," American Journal of Orthopsychiatry, 1936, 6:562-575.

[27] H. G. Gough, "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:215.

[28] E. Frenkel-Brunswik, "Mechanisms of Self-Deception," Journal of Social Psychology, 1939, 10:409-420.

that a group of teachers earned significantly more stable, dominant, extroverted and self sufficient scores on the Bernreuter scales when instructed to take the test as though they were applying for a position, than when under more neutral instructions.[29] On tests of mental ability malingerers try more items and make more errors than do intellectually inadequate persons. Malingerers also fail items that handicapped persons pass, and pass items that the defectives fail.[30] A comparative study of malingerers and authentic psychiatric cases using the Cornell Selectee Index and a shortened form of the Shipley Personality Inventory found that malingerers scored significantly higher on both tests.[31] Ruch showed that college students could fake extroversion on the Bernreuter to the extent of achieving a median at the 98th percentile on Bernreuter's norms, as contrasted with a naive median at the 50th percentile.[32] Bernreuter found that college students could produce marked shifts in their Bernreuter scores in the socially approved direction. He interpreted this finding, however, as indicating the comparative unimportance of the faking tendency. His reasoning was that had the need for giving socially approved responses operated in the first administration to any extent, the effect of special instructions to take this attitude should not have been great.[33] To Meehl and Hathaway this reasoning seemed rather tenuous, inasmuch as the occurrence of a shift merely shows that conscious and permitted faking can produce greater effects than those which may have been operating in the naive original testing.[34] Meehl and Hathaway further state:

---

[29] G. Hendrickson, "Attitudes and Interests of Teachers and Prospective Teachers," (paper read before Section Q, AAAS, Atlantic City, December 27, 1932).

[30] W. A. Hunt and H. J. Older, "Detection and Malingering through Psychometic Tests," United States Naval Medical Bulletin, 1943, 41:1318.

[31] W. A. Hunt, "The Detection of Malingering: A Further Study," United States Naval Medical Bulletin, 1946, 46:249.

[32] F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type of Personality Test," In Quinn McNemar and M. A. Merrill, Studies in Personality (New York: McGraw-Hill, 1942), pp. 229-234.

> The insignificant correlations between naive and faked scores were also used by Bernreuter to support his view, an argument which is not comprehensible in view of the gross skewness of the faked scores. What is clear from his investigation is that people are able to influence their scores to a considerable extent if they choose to, and that the average student's stereotype of what is socially desirable seems to be an individual who is dominant, self sufficient and stable.[35]

Bordin reports that students acquainted with the occupational groupings included in the Strong Vocational Interest Blank were able to simulate certain specified occupational types, even though the students were unfamiliar with the mechanics of scoring.[36] He points out that one factor determining the profile on a test of this kind is the degree of acceptance of an occupational stereotype as a self-description. It is as if a person would ask himself the question, "Who am I?" and then answer the test items in a manner consistent with the resulting self-conception. A second important factor is the degree of knowledge of the true occupational stereotype. This factor will determine the clarity of the obtained pattern.[37] Similarly, if a subject is attempting to respond as a psychoneurotic on a personality inventory, the success of the trial will be largely influenced by his understanding of the neurotic syndrome in its intimate as well as its obvious aspects.[38]

---

[33] R. G. Bernreuter, "Validity of the Personality Inventory," Personality Journal, 1933, 11:383-386.

[34] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:525-526.

[35] Ibid.

Clinical observations substantiate the above conception of malingering. Ossipov states that every malingerer is an actor who portrays an illness "as he understands it." The malingerer goes to extremes, apparently believing that the more eccentric his behavior, the more disordered he will be thought to be.

---

[36] E. S. Bordin, "A Theory of Vocational Interests as Dynamic Phenomena," Educational and Psychological Measurement, 1943, 3:57.

[37] *Ibid*., p. 54.

[38] H. G. Gough, "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:216.

In addition to exaggeration of symptoms, the malingerers tend to act as a "state" or an episode, but not a disease. For this reason, Ossipov emphasizes that the entire clinical picture must be carefully evaluated, especially the configuration of symptoms, in distinguishing a feigned from a genuine illness.[39]

Maller,[40] Metfessel,[41] Olson,[42] and Spencer[43] have studied the effects of anonymity on responses to self-rating situations and found that the requirement of signing one's name has a definite effect on scores. Kelly, Miles and Terman demonstrated the great ease with which scores on the Terman-Miles Masculinity-Femininity Test could be faked in either direction once the subjects had been informed concerning the secret of what the test measured.[44]

---

[39] V. P. Ossipov, "Malingering: The Simulation of Psychosis," Bulletin of the Menninger Clinic, 1944, 8:39-42.

[40] J. B. Maller, "The Effect of Signing One's Name," School and Society, 1930, 31:882-884.

[41] M. Metfessel, "Personality Factors in Motion Picture Writing," Journal of Social and Abnormal Psychology, 1935, 30:333-347.

[42] W. C. Olson, "The Waiver of Signature in Personal Reports," Journal of Applied Psychology, 1936, 20:442-450.

[43] D. Spencer, "Frankness of Subjects on Personality Measures," Journal of Educational Psychology, 1938, 29:26-35.

[44] E. L. Kelley, C. C. Miles and L. M. Terman, "Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality," Character and Personality, 1936, 4:206-215.

Strong,[45] Bills,[46] Steinmetz,[47] as well as Bordin,[48] have presented evidence of the ability of subjects to distort their interest patterns when taking the Strong Vocational Interest Blank.

There are several reports of MMPI simulation in the literature. Benton had nine homosexuals who were positively identified on the Mf scale retake the test and try to conceal their femininity. Six of the nine participants were able to bring their Mf scores within normal limits.[49] Meehl and Hathaway had 54 psychological trainees take the MMPI as if trying to avoid being drafted for military service, and obtained F scale T-scores of 78 or higher in 96 percent of the cases. In addition to high F scores, most of the profiles would have been clinically invalidated because of their highly unusual configurations.[50]

---

[45] E. K. Strong, Vocational Interests of Men and Women (Stanford: Stanford University Press, 1943).

[46] Marion Bills, "Selection of Casualty and Life Insurance Agents," Journal of Applied Psychology, 1941, 25:6-10.

[47] H. C. Steinmetz, "Measuring Ability to Fake Occupation Interest," Journal of Applied Psychology, 1932, 16:123-130.

[48] E. S. Bordin, "A Theory of Vocational Interests as Dynamic Phenomena," Educational and Psychological Measurement, 1943, 3:49-65.

[49] A. L. Benton, "The MMPI in Clinical Practice," Journal of Nervous and Mental Disorders, 1945, 102:416-420.

[50] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:550-551.

Meehl and Hathaway suggest that it is quite possible that in developing personality questionnaires constructed in the traditional, a priori fashion and refined by statistical manipulation, the test-maker is merely pooling sets of items to differentiate among people with respect to various test-attitude continua of little or no psychiatric relevance. The underlying disposition which leads a subject to respond in a certain way to such questions may or may not be identical with the dispositions recognized as clinical variables, nor with those that might be suggested by the item content. To Meehl and Hathaway it is quite clear on present evidence that identification cannot be established by an assumed equivalence between non-test behavior and the verbal report. Hence, both a priori selection of items and the psychological naming of a statistically homogeneous scale from its item content are fraught with possibilities of error.[51]

Guilford has explicitly called attention to the importance of the problem of test-taking attitudes as "factors" when he says:

> We must constantly remember that the response of a subject may not represent exactly what the question implies in its most obvious meaning. Subjects respond to a question as at the moment they think they are, with perhaps a lack of insight in many cases as to their real position on the question. They also respond as they would like themselves to be and as they would like others to think them to be and as they wish the examiner to think them to be. They also respond with

---

[51] *Ibid.*, pp. 553-554.

> some regard to self-consistency among their own answers. Whether these determining factors are sufficiently constant to set up individual differences which are uniform in character and so constitute common factors in themselves is difficult to say. Should any one of them be so pervasive it should introduce an additional vector in the factor analysis.[52]

Statement of the Problem

Before complete, meaningful interpretation of the M-Scales (the motivational test battery used in this study) can occur, a K validation scale must be used. Test distortion due to certain test-taking attitude continua must be examined. It was the purpose of this research project to 1) develop a K scale for the motivational test battery, 2) obtain a check on the concept of validation scale contribution in test construction, and 3) examine the problem of effects of a validity scale on test validity and reliability.

Background of Theory and Research

Among the many authors who recognize the problem of detection of malingering and falsification on objective tests there are but a few who have made specific suggestions for its solution. The inclusion of special exhortations to frankness and objectivity in the test directions themselves is common, but there is no evidence as to its effectiveness.[53] Obviously, if a subject is consciously determined to fake, he will do so; whereas if his motivation to distortion is of a more subtle, non-verbalized nature, such exhortations can hardly be expected to be efficacious. Another method is to attempt to disguise the content of items, so that the significance of a given response is less obvious. Traditional approaches to the measurement of personality render this technique practically impossible, inasmuch as the items are selected to begin with for their obvious psychological significance. Hence, unless the items are changed so greatly as to no longer elicit the desired information, they will almost inevitably continue to betray their origin. An effective use of a set of subtle items is only possible when the initial item pool is large and the initial selection of items is ruthlessly empirical. Those items whose significance would not have been guessed by the test-maker will then be equally mysterious to the testee. The presence of projective and role playing components of test-taking behavior should be recognized in objective personality inventories.[54]

---

[52] J. P. Guilford and R. B. Guilford, "Personality Factors S, E, and M and Their Measurement," Journal of Psychology, 1936, 2:118.

[53] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:527.

A A spurious anonymity using secret coding for identifying the testee is a possibility suggested by the studies cited above, but is clinically impractical and the deception involved is not desirable.[55] Lacking anonymity, it has been suggested by Olson that the name be signed at the conclusion of the test administration instead of at the top of the page.[56] This suggestion was carried into practice by Maller in his Character Sketches.[57] In addition, he also stated the questions in the third person (indirect) form, requiring the subject to indicate whether he was the same or different from the person described. Maller presents evidence that this procedure aroused considerable less annoyance in his subjects; however, direct proof that this decrease in annoyance led to increased validity is lacking. Meehl expresses doubt whether or not the removal of personal reference is wholly desirable because there is reason for believing that the same role playing and self-deception which operate to invalidate some of the measurements are an important factor in making other measurements possible.[58]

---

[54] P. E. Meehl, "The Dynamics of Structured Personality Tests," Journal of Clinical Psychology, 1945, 1:296-303.

Another technique for reducing the effect of signing one's name is to have the items printed on cards which are then sorted by the subject. Such a procedure makes all writing unnecessary and it is assumed that the feeling that one is making a permanent record of his personal failings is lessened. This method has been employed by Maller in a revised test (Personality Sketches) and by Hathaway and McKinley in the MMPI.[59,60] However, evidence supporting the above assumption concerning the increased validity of performance is lacking.

---

[57] J. B. Maller, Character Sketches (New York: Bureau of Publications, Teachers College, Columbia University, 1932).

[58] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:527-528.

Although all of these strategems may have a considerable value particularly in the aggregate, the fact still remains that they do not by any means remove the possibility of faking. What is much more important, they are mainly directed at the sort of conscious falsehood which most investigators have stressed, while ignoring the more subtle tendencies to self-deception which are probably of even greater importance in affecting scores. Also, they neglect to stress the existence of trends in the opposite direction--namely those trends which exaggerate the apparent abnormality or maladjustment of the individual. Meehl and Hathaway state that it is only natural that the tendency of a testee to put himself in a favorable light should have received more attention than the contrary tendency. However, there is considerable evidence that this latter tendency does exist and that it is a

---

[59] J. B. Maller, "Personality Tests." In J. M. Hunt, Personality and the Behavior Disorders (New York: Ronald Press, 1944).

[60] S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule: I. Construction of the Schedule," Journal of Psychology, 1940, 10:249-254.

much more important factor in determining scores on inventories than has generally been supposed.[61]

It is also probable that certain systematic differences in item-interpretation, not necessarily a function of personality dynamics of the defensive or self-critical sort but relatively neutral psychologically (semantic variation), lead to score deviations which are misleading. Such problems have been investigated by Benton,[62] Eisenberg,[63] and Eisenberg and Wesman.[64]

A more fruitful attitude was taken by Rosenzweig in which he reiterated the fact of untrustworthiness of self-ratings and indicated that instead of trying to eliminate completely these sources of error, the test-maker should recognize them and attempt to correct for them in interpreting the results. He says:

> Astute phraseology in the instructions and questions of the test have sometimes been resorted to, but such expedients are rarely very effective. Might it not be more effective

---

[61] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:528-529.

[62] A. L. Benton, "The Interpretation of Questionnaire Items in a Personality Inventory," Archives of Psychology, 1935, Number 190.

[63] P. Eisenberg, "Individual Interpretation of Psychoneurotic Inventory Items," Journal of Genetic Psychology, 1941, 25:19-40.

[64] P. Eisenberg and A. Wesman, "A Consistency in Responses and Logical Interpretation of Psychoneurotic Inventory Items," Journal of Educational Psychology, 1941, 32:321-338.

> to recognize at the outset that such tests have certain limitations that can never be completely circumvented and then go on to the measurement of these limiting factors themselves, thus obtaining information by which a correction may be applied to the subject's answers?[65]

Rosenzweig's specific proposal for achieving this end was to include among the usual self-rating items a set of items of the form "I should like to be the sort of man who ...," on the theory that if the test-maker knew something of the strength of certain "ideal-self" trends in the person, the investigator could make appropriate correction for these trends in interpreting responses to the traditional items. Rosenzweig, however, never carried this idea into practice. On the other hand, Meehl and Hathaway consider that this approach would be relatively ineffective, since they feel what is desired is not a statement of the strength or number of ideals for the self, but a measure of the extent to which they are allowed to distort responses. In other words, a subject might easily have lofty ideals verbally expressed, but might be too honest, insightful, objective or self critical to distort his responses into agreement with these ideals.[66]

---

[65] Saul Rosenzweig, "A Suggestion for Making Verbal Personality Tests More Valid," Psychological Review, 1934, 41:400-401.

[66] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:529.

Maller attempted to solve this problem in another way in his Character Sketches by including a small set of items which were supposed to measure the subject's "readiness to confide." The occurrence of normal, well-adjusted scores in combinations with a low-measured "readiness to confide" would lead one to be skeptical of the validity of the measurement.[67] However, the "readiness to confide" items were themselves self-ratings on readiness. In the later form called Personality Sketches, Maller does not make use of the "readiness to confide" concept, so it may be assumed that it was unsuccessful or at least did not materially improve validity.

Meehl and Hathaway, carrying Rosenzweig's thinking to its logical conclusion, consider the obvious procedure to follow is to give the subject a good chance to distort his answers in accordance with some self picture or conscious facade, and observe the extent to which he does so.[68] Of course, the difficulty here is that such a procedure requires a knowledge of the objective and subjective facts which are usually inaccessible. Here there are apparently three possibilities open to the test builder. First, he may sidestep the problem of getting directly at the objective truth, and attempt to establish falsehood by obtaining internal contradictions, a technique employed by Maller in his earlier test. Cady, in his application of a modified form of the Woodworth Psychoneurotic Inventory to the measurement of juvenile incorrigibility, had earlier made use of repeated items to increase reliability of the scores (although the aim of detecting inconsistency of the fake sort was not explicit in his rationale).[69] Each question appeared twice, once in each section of the test, except that in the second appearance the question was phrased in the negative. Theoretically the subject's response should also be reversed, and the number of failures to reverse is an indication of some inconsistency. Hence, a measure of non-cooperation or dishonesty would be obtained. The inconsistency score obtained in this way was to be subtracted from the adjustment score to get a sort of corrected score as proposed by Rosenzweig. However, Meehl and Hathaway point out that it is by no means obvious that the shift to a negative form of item will leave the projective properties of the stimulus simply reversed in meaning; so that the fact of an inconsistency in the strict logical sense would not necessarily imply lack of cooperation or dishonesty. However, it would seem reasonable that a very large number of such inconsistent pairs would cast grave suspicion upon the scores, either for dishonesty or some equally serious reason.[70] This technique also was abandoned by Maller in his revised instrument.

---

[67] J. B. Maller, Character Sketches (New York: Bureau of Publications, Teachers College, Columbia University, 1932).

[68] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:530.

---

[69] V. M. Cady, "The Estimation of Juvenile Incorrigibility," Journal of Delinquency Monographs, 1923, Number 2.

The second method of using distortion is to present opportunities for answering in an extremely favorable way, but in a way which could almost certainly not be true. This idea was employed by Hartshorne and May in the Character Education Inquiry.[71] If it is assumed that there are very few aspects of behavior for which one could have complete confidence that no subject would be "ideal" in them, it is necessary to present a considerable number of such opportunities and progressively reduce the probability that any individual would be as described. In this sense, everyone would possess at least a few highly desirable traits, and no one would be the possessor of all. Without knowing anything whatsoever about a particular person, the test-maker could write on rational grounds a list of extremely good and rare human qualities which is statistically absurd to suppose will all or in large part be possessed by the individual. If the testee says, however, that he has all or a great many of them, it can be decided that he is not telling the truth. The answers to these items could yield strong evidence for deception.[72]

---

[70] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:530.

[71] H. Hartshorne and M. A. May, Studies in Deceit (New York: Macmillan, 1928).

The Humm-Wadsworth Temperament Scale has made use of the socially-desirable response method.[73] Humm and Wadsworth deserve credit for having been among the first investigators of structured personality measurement to lay great stress upon the problem of detecting non-cooperation and distortion of response when evaluating a particular profile of scores. They were also among the first to adopt an explicit and uncompromising empiricism in selecting items from a large initial pool. The two scales which serve as "checks" or "correctors" for the remainder of the profile on the Humm-Wadsworth are the "normal" component and the "no-count." The "normal" component attempts to assess the strength of a general inhibiting, controlling or normalizing factor in personality which Humm and Wadsworth considered always present to act as a "brake" upon strong abnormal tendencies on the other variables. This means that in interpreting a given profile the significance of any deviation on one of the abnormal components must be established with the size of the normal score in mind. However, Meehl and Hathaway question Humm and Wadsworth's claim for the normal component.[74]

---

[72] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:531.

[73] D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Scale," American Journal of Psychiatry, 1935, 92:163-200.

The "no-count" is based upon the number of items to which the subject responds in the negative. Inasmuch as approximately 76 percent of the scored items of the Humm-Wadsworth are "obviously" suggestive of abnormality when replied to affirmatively, the "no-count" is to some extent a measure of the testee's tendency to avoid, consciously or otherwise, saying "bad" things about himself when taking the test. That this relationship occurs is further supported by the tendency for the no-count to correlate positively (.77) with the normal component and negatively with the various abnormal components.[75] If the no-count is excessively great, the inference is that the subject has responded in a very defensive or possibly stereotyped fashion; and therefore the particular testing is of doubtful validity. Humm and Wadsworth state that as high as 25 to 30 percent of normals seem to invalidate their scores in this way, a proportion which seems to Meehl and Hathaway to be impractically high for clinical purposes.[76] Later, Humm, Storment and Iorns attempted to reduce the proportion of useless tests by a "correction" for the no-count based upon multiple regression procedures.[77] A study of hospitalized psychiatric cases by Arnold indicated that even the exclusion of cases with invalid no-count did not result in any greater validity clinically than was obtained using all cases.[78] Humm stated that improved multiple regression techniques have resulted in a marked reduction in the proportion of test misses and of uninterpretable profiles.[79]

---

[74] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:531.

[75] D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Scale," American Journal of Psychiatry, 1935, 92:174.

Washburne, in revising his Test of Social Adjustment, included a set of 21 items modeled after the "lie" items of Hartshorne and May and referred to the total score on this set as "objectivity." This score was included to detect both lying and unintentional inaccuracy. An extremely low objectivity score was said to invalidate the test as a whole. A weighted objectivity score was included in the total score on the entire test.[80]

---

[76] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:531.

[77] D. G. Humm, R. C. Storment and M. E. Iorns, "Combination Scores for the Humm-Wadsworth Temperament Scale," Journal of Psychology, 1939, 7:227-253.

[78] D. A. Arnold, "The Clinical Validity of the Humm-Wadsworth Temperament Scale in Psychiatric Diagnosis," (unpublished Doctor's thesis, University of Minnesota, Minneapolis, 1942).

[79] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:532.

The third technique available is the empirical derivation of a fake scale by making use of the item shifts obtained when persons take a test under normal naive conditions and then are retested with instructions to fake. This method has been used by Ruch to construct an "honesty" key for the Bernreuter.[81] To Meehl and Hathaway it is interesting that such a procedure "so logical and straightforward, invented to solve a problem so obvious and insistent, should have been employed for the first time over twenty years after the appearance of the first personality inventory."[82] Ruch says:

> The argument is rather simple. If answers to items on a test like the Bernreuter can be faked at all, the chances are that some are easier to fake than others. Therefore, it should be possible to give each item a weight to represent the extent to which it can be faked by the average college student. This was done by tabulating the frequency of each answer to each question for the standard condition and for the influenced condition. These frequencies were converted into percentages, and an honesty weight was assigned to each reply according to the magnitude of the critical ratio

---

[80] J. N. Washburne, "A Test of Social Adjustment," Journal of Applied Psychology, 1935, 19:125-244.

[81] F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type of Personality Test." In Quinn McNemar and M. A. Merrill, Studies of Personality (New York: McGraw-Hill, 1942), pp. 229-234.

[82] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:532-533.

> of the difference between the frequency of the reply in the honest and in the influenced condition.[83]

Ruch seems to have been the first investigator to attempt empirical derivation of a fake key for a question-answer personality inventory.[84]

As was stated earlier by Meehl and Hathaway there is evidence of a tendency on the part of some testees to make themselves appear in a "bad" light in taking personality tests. Such a tendency is difficult to characterize because it may occur on several different bases. A patient in the hospital may engage in a type of malingering for strictly conscious reasons, presenting a profile on a test which shows abnormalities out of all reasonable proportion to what is apparent from other considerations. Again, there may be somewhat general traits of verbal pessimism or self-deprecation which act to distort systematically the results of personality measurement. Meehl and Hathaway have dichotomized the test-attitude continuum by the two opposed terms defensiveness and plus-getting. However, they make no implication concerning the degree of conscious, deliberate deception involved in either. The corresponding extremes of deliberate deception are referred to as faking good and faking bad, respectively. It was recognized that, like the defensive tendency, the plus-getting tendency might exist in all degrees from a mild self-criticality or merely objectivity to a deliberate, conscious attempt to make oneself look abnormal. Whether this represents simply the extreme of a continuum with faking good at the opposite end, or an entirely new and different factor, was undecided by these authors.

---

[83] Ibid., p. 231.

[84] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:533.

Meehl and Hathaway state:

> In any case it would be desirable to develop a scale for detecting these test-taking tendencies to put oneself in a bad light when answering a personality inventory, so that allowance might be made in such cases in the light of a deviant score obtained on such a scale.[85]

Summary Statement of Organization

The remainder of this investigation is presented in four sub-divisions: in Chapter II previous attempts to develop validity keys are epitomized; the motivational research study with its instrumentation and the procedures used in the present investigation are discussed in Chapter III; in Chapter IV, K Scale development, sex differences in K item selection, K Scale reliability and validation are discussed; and in Chapter V the summary and conclusions of the investigation are presented.

---

[85] *Ibid.*, pp. 533-534.

CHAPTER II

PREVIOUS ATTEMPTS TO DEVELOP VALIDITY KEYS: A REVIEW OF THE LITERATURE

For the present investigation all of the major validity scales were reviewed. The order of discussion of the various keys was based upon the chronology of their development with the earliest devised scales discussed first. In addition to the major scales, however, several minor scales had to be considered in order to bring about a more lucid understanding of the major keys.

The L Scale

The "lie" scale of the MMPI attempts to identify those individuals who try to falsify their score by choosing responses which they feel are most acceptable socially.[1] The original fifteen MMPI items making up the L scale were selected under the inspiration of the work of Hartshorne and May.[2] Each of the items presents a situation desirable socially, but which is rarely true of an individual. It was recognized by Hathaway and McKinley that extremely conscientious persons would frequently have more than the average of the L items validly positive, but it was assumed that for a person to have six or eight such items marked was highly improbable.[3] It was concluded by these investigators that the fifteen items of this type scattered among the main body of the items constituted a fairly subtle trap for anyone who wanted to give an unusually good impression of himself.

---

[1] S. R. Hathaway and J. C. McKinley, Manual for the MMPI (New York: The Psychological Corporation, 1945).

[2] H. Hartshorne and M. A. May, Studies in Deceit (New York: Macmillan, 1928).

The standardization procedure revealed that among the various normal groups the mean score on the L items lay between three and five items. The frequency curves were all skewed sharply in the positive direction. Few individuals obtained raw scores of seven or more. Only two or three percent exceeded ten items. These values were arbitrarily called the 60 and 70 T-score points, respectively. As more data were accumulated Hathaway and McKinley concluded that the original tentative assumptions regarding the meaning of L were in the main correct, but other valid interpretations of L in the range from T-score 56-70 also existed. The original arbitrary assignment of T-scores had been too conservative, and more emphasis was placed on the T-score range of 56-60. To Hathaway and McKinley the positive presence of the rise in the L score seemed quite valid as an indicator that the individual taking the test was being dishonest and might be somewhat unreliable. However, if no rise in L was observed, these investigators offered no positive or clear interpretation.[4]

---

[3] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:537-538.

To check the assumption that L would not identify the more sophisticated subject, an experiment was performed with 53 male psychology students. The participants, who had completed a considerable portion of their training in psychology, were asked to take the MMPI twice. The first administration was done in the standard manner. In the second administration the group was asked to make certain that they would be acceptable to army induction. Half the group took it with fake good instructions first, half second. Through this procedure a faked good record and a normal record were both obtained.[5]

These records showed no appreciable rise in L. It is also true, however, that the majority of the profiles were only slightly better than the corresponding non-faked profiles.

Although one might conclude from the design of the above experiment that the outcome simply tested the participants' willingness and/or ability to fake good, Hathaway and McKinley held that the results demonstrated that the intent to deceive is not often detectable by L when the subjects are relatively normal and sophisticated.[6]

---

[4] __Ibid__.

[5] __Ibid__.

[6] __Ibid__.

Cottle, however, found additional support for one of the conclusions of Hathaway and McKinley, namely, that sophisticated, bright individuals tend to score low on L items. In his study with 100 high level college students on the MMPI Cottle found that the mean L score on the card form was 2.54 raw score points and for the booklet form was 2.73 raw score points.[7] However, no conclusions can be reached from this study concerning the hypothesis that the L scale is a valid key differentiating between individuals who wish to fake good and those who do not.

Hovey discusses three cases of individuals who discovered the scoring purpose of cuts on MMPI cards. He says that the L score in these cases was zero.[8] On the other hand, Cofer and others when studying the effect of malingering on the MMPI found that the instructions to fake a normal profile raised L scores.[9]

From the evidence presented in the literature no definite conclusions can be drawn at this time concerning the efficacy of the L scale. No probability statement regarding its differentiating power is offered. The literature only rationally suggests that factors of degree of training and psychological sophistication may influence unduly an individual's performance on L items.

---

[7] W. C. Cottle, "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, 1950, 34:255-259.

[8] H. B. Hovey, "Detection of Circumvention in the MMPI," Journal of Clinical Psychology, 1948, 4:97.

[9] C. N. Cofer, J. Chance and A. J. Judson, "A Study of Malingering on the MMPI," Journal of Psychology, 1949, 27:491-499.

The F Scale

In the original publication of the MMPI the F scale was not presented as an empirically validated variable. Its validity was assumed on a priori grounds. The key was composed of 64 items which were selected because they were answered with relatively low frequency (10 percent or less) in either the true or false direction by the main normal group. The scored direction of response was the one which was rarely made by unselected normals. Additionally, the items were chosen to include a variety of content so that it was unlikely that any particular pattern would cause an individual to answer many of the items in the unusual direction. The relative success of this selection of items, with the deliberate intent of forcing the average number of items answered in an unusual direction downward, was illustrated in the fact that the mean score on the 64 items ran between two and four points for all normal groups. The distribution curve was extremely skewed; the higher F scores approached half the total number of F items. In distributions of normal persons the frequency of scores dropped rapidly at about seven items and was at the two or three percent level of chance by a score of twelve. Because of this quick cutting off of the curve and scores seven and twelve were arbitrarily assigned T-score values of 60 and 70 in the original F table.[10]

From the first Hathaway and McKinley recognized the F represented several interpretations. 1) The subject would need to sort almost all of the items according to expectation in order for these low scores to result, and any error in recording, such as mistaking true items for false items and the like would raise the F score appreciably. 2) If a subject could not understand what he was reading adequately enough to comprehend fully the answers to these items, the F score would obviously be higher. 3) Persons who were highly individualistic and independent might honestly make infrequent responses to F score items. For example, such individuals might admit to disliking children and not believing their mother was a good woman. 4) It was early discovered that schizoid subjects and subjects who apparently wished to put themselves in a bad light also obtained high scores. Meehl and Hathaway felt that the schizoid group obtained high scores because they said unusual things due to delusional or other aberrant mental states in responding to the items. This was referred to as distortion since it was considered that an impartial study would not justify the patient's placement. Among more normal persons some high scores were also observed where the individual had rather unusual ways of responding to conventional stimuli. For example, to the item, "I have had periods in which I carried on activities without knowing later what I had been doing" most persons answered false. Some individuals, however, included periods of sleep in the implication of the item. One might argue that such ways of thinking are often allied to schizoid mentation generally and that the answers in this case indicate a true abnormality. At the very least, however, the person is responding to some items in a way that differs from that of most individuals. Meehl and Hathaway conclude that such persons might, therefore, not be appropriately approached through this method of personality measurement. To them it seemed reasonable that there are individuals whose habitual ways of reacting to items are so different from other persons that measurement of their personalities through the use of verbal items of this type would reflect the unusualness of their reactions to the items more than any clinical abnormality.[11]

---

[10] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:535-536.

Clinical experience suggested to Meehl and Hathaway that the usual critical score of T equals 70 was too low in the case of F. They found that scores ranging up to T scores of 80 were often more a reflection of validly unusual symptoms and attitudes than an indication of invalidity in the rest of the profile due to misunderstanding, etc. However, scores above this strongly suggested an invalid record.[12] As a result, it was decided that scores above 70 would indicate the whole record to be invalid, except in the special cases mentioned above of schizoid tendencies. Scores from 60-70 would be considered opened to suspicion; scores from 50-60 would be considered a reliable sign that item comprehension, clerical work, etc., had been satisfactory and that the subject was similar to persons in general.

---

[11] *Ibid.*, p. 536.

When the MMPI was administered to incoming servicemen it was possible to consider the F score as evidence of an attempt to malinger and to obtain fallaciously bad scores on the other MMPI scales. To check on this interpretation, a similar study to the investigation conducted on the L scale (see pages 25-29) was devised. A group of 54 service men who had completed a considerable portion of their training in psychology were asked to take the MMPI twice. The group took the MMPI in the standard way and also took it under instructions to assume that they wished to avoid being accepted in the draft; and in order to be rejected they were to obtain adverse scores without giving themselves away. The order was reversed for half of the group for the test administration. Through this plan a faked bad record and a normal record were obtained. The data reveals that 96 percent of the faked bad records had a raw F score of 15 or more (T-score of 78 or greater). The researchers concluded that even these men who were somewhat cognizant of psychological measurements betrayed themselves when they attempted to fake a bad record.[13] The F scale was a good device for identifying the intentional faking that could be set up in an experimental situation.

---

[12] _Ibid._, pp. 536-537.

Kazan and Sheinberg found, however, that a high F score is rarely an invalidating factor with abnormal subjects. In their investigation of 170 maladjusted, male service men, all the items of the F score were not answered in the infrequent direction less than 10 percent of the time by normals, and that the percentage was but little higher for miscellaneous abnormal subjects.[14] Schmidt, likewise, found that the profiles for psychotics rose more sharply at F than for any other clinical group.[15] Schneck also found a high F score less valid in a study of character disorders in an army disciplinary barracks.[16] Cofer and others show that subjects attempting to fake emotional upset were detected easily by the F score.[17] Another study of faking on the MMPI by Gough reports that feigned psychotic curves can be detected by being too low on neurotic scales, too high on psychotic scales, and by a significantly elevated F scale. Gough also found that feigned neurotic curves were identified by high F and low K scores.[18] Mechanical sorting using an index of F minus K correctly selected 82 percent of these feigned neurotic profiles. This study used eleven individuals with a background of psychology or psychiatry as contrasted with controls of thirteen hospitalized paranoid schizophrenics and 57 severe psychoneurotics. Gough concludes that relatively skilled persons are unable to simulate either a psychoneurotic or psychotic condition on the MMPI in such a way as to avoid detection. Hunt used a group of 109 psychology students and 74 Navy general court martial prisoners to investigate the effect of deliberate deception.[19] He substantiated Gough's discovery that an index of F minus K correctly identifies a substantial proportion of malingered or faked abnormal profiles. However, Hunt concluded that the index was of no use in detection of faked normal profiles.

---

[13] *Ibid.*, p. 537.

[14] A. T. Kazan and I. M. Sheinberg, "Clinical Note on the Significance of the Validity Score F in the MMPI," *American Journal of Psychiatry*, 1945, 102:181-183.

[15] H. O. Schmidt, "Test Profiles as a Diagnostic Aid: the MMPI," *Journal of Applied Psychology*, 1945, 29:115-131.

[16] J. M. Schneck, "Clinical Evaluation of the F Scale on the MMPI," *American Journal of Psychiatry*, 1948, 104:440-442.

---

[17] C. N. Cofer, J. Chance and A. J. Judson, "A Study of Malingering on the MMPI," Journal of Psychology, 1949, 27:491-499.

[18] H. G. Gough, "Simulated Patterns on the MMPI," Journal of Abnormal and Social Psychology, 1947, 42:215-225.

[19] H. F. Hunt, "The Effect of Deliberate Deception on MMPI Performance," Journal of Consulting Psychology, 1948, 12:396-402.

A later report on the F minus K index by Gough suggests that the sampling distribution of F minus K is reasonably normal and the index is not particularly distorted by psychiatric abnormality.[20] He presents data from Sweetland which support the use of F minus K.[21]

After reviewing the MMPI literature Cottle found that high scores on F are caused by carelessness of the subject, scoring errors, psychotic state, or deliberate faking of an abnormal profile. He, too, concluded that the F scale is of little use in detecting faked normal profiles.[22]

From the evidence presented in the literature Cottle's conclusions concerning the F scale appear rationally sound. However, as in the case of the L scale, no translucent probability statement is offered regarding the differentiating power of F in distinguishing between adequate and inadequate performance (as outlined by Hathaway and McKinley on pages 30-33 above).

---

[20] H. G. Gough, "The F Minus K Dissimulation Index for the MMPI," Journal of Consulting Psychology, 1950, 14:408-413.

[21] A. Sweetland, "Hypnotic Neurosis--Hypochondriasis and Depression," Journal of Genetic Psychology, 1948, 39:19-105.

[22] W. C. Cottle, "The MMPI: A Review," Kansas Studies in Education, 1953, 3:6-9.

The K Scale

Meehl and Hathaway conceptualized two approaches to the problem of identifying the attitude a subject takes toward the items in a personality inventory. First, the investigator may have the subject deliberately assume a generally defined attitude. For example, faking might be directed toward obtaining either adverse or desirable scores. A "normal" set of responses must be obtained relatively simultaneously with the faked responses if a reference point is to be determined. The faked and normal records can then be contrasted for study to discover the items which are most frequently changed from the normal records as contrasted to the fake records. Secondly, the investigator may choose records in which there is presumptive likelihood that a special attitude has been assumed. As in the first approach a "normal" set of responses must be obtained simultaneously for comparison.[23]

Using the direction to fake approach several scales were derived by Meehl and Hathaway. The good and bad fake scales were found to be composed of different sets of items. Either of the procedures provided a scale that would be about as good for the other type of faking as it was for the one from which it was derived when such scales were applied to test cases not used in the original derivations. Using two such scales separately did not materially increase the accuracy of prediction.

---

[23] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:538-539.

In the second line of experimental approach Meehl and Hathaway also derived several subdivisions. Among presumably functional and normal records cases were often identified which were so abnormal as to indicate that the individual should have been hospitalized. The investigators attempted to discover items which would differentiate normal from clinically diagnosed abnormal persons. For the counterpart to this approach they also selected for item analysis hospitalized cases whose records showed a normal profile. Using this approach Meehl and Hathaway experimentally derived four scales.

Derivation of the L6 Scale. The most important finding of the investigators was that whichever of the methods used, as was with the case with the faked approach above, the resultant scales were about equally effective. These scales were also fairly effective in differentiating the fake group as well. After two years of this experimentation all of the promising scales were cross-validated on a new sample. A single best scale was derived which was originally called L6.[24]

---

[24] Ibid., pp. 539-540.

L6 was derived by an item analysis of the responses of 25 males and 25 females in a psychopathic hospital. These subjects' MMPI profiles showed an L scale ("lie" key) of T equal 60 or more. In addition, these individuals were predicted to obtain abnormal profiles because of the clinical diagnoses given to these cases by the psychiatric staff. However, the scores on the MMPI profiles fell within the normal range.

Two restrictions were employed in the selection of the criterion group. All of the individuals were characterized by deviant behavior; however, they obtained relatively normal profiles and were termed "misses" for the MMPI. In addition, all criterion cases were characterized by having a tendency to obtain elevated scores on the L scale.

The item responses of these 50 cases, analyzed separately for males and females, were compared to item frequencies from previous standardization groups. In all, 22 items were chosen as a result of this comparison on the basis of a 30 percent discrepancy between validation and standardization groups.

Because the criterion group was assumed to desire good scores, the larger raw scores on these items were in the same direction as the larger raw scores on the L scale. The item content suggested an attitude of denying worries, inferiority feelings, and psychiatrically unhealthy symptoms, together with a disposition to see only good in others as well as oneself.

Cross-validation of L6. Following the final choice of L6 as the best of the scales available, Meehl and Hathaway subjected the validity key to more careful study. Hospital and normal records were examined to discover whether L6 would be helpful in interpreting individual profiles. Relatively few data were found on normal cases, but on hospital cases a fairly extensive symtomatic summary was available. By examining the profile for normalcy it was determined whether or not the L6 deviated in an upward direction. It was assumed that an upward direction indicated that the patient had attempted to place himself in a good light. As a result of this study L6 was judged effective but left much to be desired.[25]

The L6 scale as a measure of defensive and plus-getting attitudes. To the investigators L6 appeared as adequate for the detection of plus-getting as was N (see section below) or any of the other experimental scales. Accordingly, the records of a new series of presumable normal persons showing deviant profiles was examined. The L6 scale again appeared to work at the plus-getting end of the test-attitude continuum. That is to say, a relatively low score on L6 could be used to underinterpret an otherwise deviant profile and thus avoid some of the presumable false positives in the normal population sample. Thus L6 seemed useful at both ends of the test-attitude continuum: defensiveness and plus-getting.

---

[25] Ibid., pp. 540-542.

Refinement of the L6 scale. The most outstanding difficulty in the above procedure was that L6 tended to be low on severe depressive or schizophrenic patient records. This led to an underinterpretation in spite of the fact that the patients were grossly abnormal. To correct for the underinterpretation tendency, items were added that would work in the opposite direction. To choose these items Meehl and Hathaway studied the item tabulations for the groups of psychological trainees above who had attempted to fake good and bad scores. In the above study there were many items which showed no tendency to change with an alteration in the test-taking attitude. The percent of true or false remained constant whether the attitude was the normal one or the faked one. From among these items, a sub-group was chosen which showed differences between schizophrenic and depressive criterion groups and general population normals. Meehl and Hathaway admit that the procedure rested upon the insecure assumption that any item that did not appear to be affected by the test-taking attitude (as approached by a normal person attempting to fake good or bad, but occurred as a frequent item to differentiate depressed or schizophrenic patients) would be useful in correcting the tendency of L6 to go too low for schizophrenic and depressed patients. Such an item was scored in a way that would make it work against the tendency of the L6 scale. Eight items were selected by this method. The effect of adding these eight items to the 22 on L6 was to elevate slightly the mean score of normals and to make it more nearly approach the mean score of abnormal cases on the complex of all 30 items.[26]

Derivation of the K scale. As a final step in the refinement of the L6 scale the above eight items were combined with the 22 L6 items into a single scale which was designated K. The K scale represents the final outcome of many experiments in the field of measuring test attitude. Meehl and Hathaway state:

> The K scale is far from perfect for its purpose as measured by the various available data. Generally speaking it is about as good as any other single scale yet derived. In individual applications it is inferior now to one scale and now to another but the differences are never great enough to be very significant practically and the small number of items in this scale gives it a distinct advantage over one or two of the longer scales such as N.[27]

Because the K scale was derived as a correction scale for improving the discrimination yielded on the already existent MMPI scales, it was not assumed to be measuring anything which in itself was of psychiatric significance.

Meehl and Hathaway considered that it was first necessary to choose criterion cases of the sort on which K could conceivably be of value. It was apparent to these investigators that such cases would be characterized by the presence of what was called borderline profiles, that is, individuals who possessed T-scores between 65-80. In studying hundreds of deviant profiles after the addition of K, almost no individuals were found with T-scores above 80 in the normal sample; and it was not statistically profitable to correct elevations of such magnitude to the point of calling them normal. On the other hand, when a curve showed no elevation at all above 65, even the presence of a high K score did not enable the examiner to form any adequate notion of what the peak would be had the K factor not been operating to distort the results. There were apparently upper and lower limits beyond which deviations on K could not effectively operate. Profiles showing subtest scores above 80 were interpreted as probably abnormal no matter how low K fell. If a profile showed no subtest scores above 65 it was unknown whether a high K meant the profile should be adjusted toward more severe scores or was merely that of an actually normal person who for some reason took a defensive attitude when being tested.

---

[26] Ibid., p. 543.

[27] Ibid., pp. 543-544.

Validation of the K scale. Meehl and Hathaway judged that the kind of curve which gave interpretative difficulty, and which could be improved by knowledge of the influence of K, would be a curve in the doubtful, borderline region. Accordingly, a group of cases from the normal and hospital groups was chosen on the basis of having achieved such borderline curves. For this study all cases in the files showing at least one personality component elevated as high as T equal 65, but with no component elevated to T greater than 80. Among the normals, there were 71 males and 103 females having such borderline curves. Corresponding to these cases, 129 abnormal males and 208 abnormal females were located with similar borderline profiles. The data for the two sexes were treated separately.

The analysis of these data was in terms of the ability of the K scale, used mechanically, to separate the curves of the actual normals from those of the actual abnormals. The procedure was to arrange the whole set (normals and abnormals combined) for each sex in order of the magnitude of their scores. The distribution of K was cut on the basis of the proportion of normals and abnormals in the sample, calling all cases above the cut abnormal and all those below normal. Setting up a fourfold table on this basis, a chi-square of 20.436 for the males and 29.540 for the females was obtained. Both of these were highly significant (P equal to less than .0010) with one degree of freedom. If instead of locating an optimal cutting score, the K distribution was cut at the mean of the general population K distribution (T equal 50), the cutting point of the males was unchanged. However, the cutting point for the females shifted enough to lower their chi-square to 17.750, which was still highly significant. If one considered miscellaneous profiles which lie in the borderline range between 65 and 80, regardless of the kind of elevation and irrespective of the clinical diagnosis of those who were clinically abnormal, one could separate them into actual normals and abnormals significantly better than chance by using a cutting score on K. In this instance Meehl and Hathaway emphasized that K was operating chiefly as a suppressor of certain test-taking tendencies, for K by itself did not differentiate unselected normal and abnormal cases. In terms of percentages, it was found that for the males, 72 percent of the abnormals and 61 percent of the actual normals were correctly identified. For the females, 66 percent of the abnormals and 59 percent of the normals were correctly classified. These percentages were based upon the separations with K equal 50, taking no account of the actual normal-abnormal proportions among the above cases.[28]

Refinement of the K scale. Evidence from examination of the test "misses" disclosed by K in the data, combined with knowledge of the correlation between K and other MMPI scales, indicated that the K correction was more important in some scales than in others. Therefore, it was decided to analyze the borderline groups in terms of the peak elevation of their profiles in the attempt to identify those particular curves on which K could be used with profit.

---

[28] *Ibid*., p. 545.

The entire group of 511 borderline curves (males and females, normals and abnormals pooled) was divided into eight sub-groups, each sub-group composed of cases having the peak score on the same one of the eight MMPI personality components.

The normals and abnormals having borderline curves with the same peak score were then separated mechanically by the use of a cutting score on K. The proportion of cases above the cutting score was determined on the basis of the proportion of actual abnormals versus normals in each sub-group. Meehl and Hathaway state that it was unavoidable in the analysis because the relative proportions of actual normals and abnormals varied widely from scale to scale and the use of the mean of K would have been grossly misleading. For the eight groups studied in this manner, only three showed a significant chi-square between the 10 percent and 20 percent level of significance. It seemed, therefore, that the K factor could be used with profit in interpreting some kinds of profiles but not others. The failure to discriminate with K when grouping profiles by peak score did not establish that a K correction might not be profitably added to the single scores themselves.[29]

---

[29] *Ibid*., p. 546.

Cross-validation of the K scale. One other validating study was done of K in the original investigation. A group of 22 normals and 22 abnormals who were employed in a previous study was used.[30] The normals consisted of a random selection from a large group of profiles showing any elevation of 70 or over. The abnormals consisted of a heterogeneous group also having at least one subtest score over 70. All cases were chosen randomly from hospital cases. Because these groups had been selected for a different investigation, they had not entered into the derivation of K in any way. Without regard for any other information concerning the profiles, all cases showing K greater than 50 were called normals. The cutting score was also independent of the statistics of the original group. To Meehl and Hathaway the K scale worked phenomenally well here. Of the entire group of 44 cases, 37 were correctly classified when using K in this way, a total of 85 percent "hits." Here it was the purpose to separate normals and abnormals all of whom possessed deviant profiles. The investigators concluded that this percent was quite impressive considering the task set for K. Of the seven errors in classifying, six were false positives (cases of normals showing elevated profiles and K greater than 50 and termed abnormal). The chi-square for the fourfold table of these data was 21.569 which with one degree of freedom was highly significant (P less than .0001). Meehl and Hathaway conclude:

---

[30] P. E. Meehl, "An Investigation of General Normality Control Factor in Personality Testing," Psychological Monographs, 1945, 59:Number 4.

> Here we have striking evidence of the validity of K when used to differentiate between deviant curves of actual normals and abnormals. We are not prepared to explain the superiority of this result to that originally encountered, except to say that the range of abnormal scores in the present analysis was from 70-90 whereas in the original analysis borderline scores were defined as lying between 65 and 80. In what way this could make K appear to function more effectively in the one case than the other is not clear. Also the present study involved only males where K in general seems to work a little better than on females.[31]

Discrepancy in the efficacy of K. The fact that K was less effective when applied to some scales than others would suggest separate interpretations or cutting scores. Furthermore, the classification into normal and abnormal on the basis of a single arbitrary cutting score obviously sacrifices some quantitative information about the actual magnitude of the K score. Meehl and Hathaway, however, did not tend to propose such a cutting method as the most efficient manner of application for K. They simply used that form to indicate that K possessed differentiating power for what it was hoped to differentiate.

With the exception of Hy and D clinical scales the correlations of K with the other MMPI variables were consistently negative. This was to be expected if K represented the defensive, lying, or self-deceptive test-taking attitude it was derived to measure. It might be thought that such low correlations as occur in Table I below would preclude any possibility of the use of K as a suppressor. However, there is a tendency for the scales on which K seems valid by the chi-square test to show higher correlations. But for the use to which K was put, correlations as low as .20 were utilized to yield very significant and useful improvements in discrimination.[32]

---

[31] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:546-547.

Table I

Intercorrelations of K with Other MMPI Variables

	Hs	D	Hy	Pd	Pa	Pt	Sc	Ma
Normal males	-.30	-.15	.48	-.17	-.07	-.67	-.59	-.36
Normal females	-.35	-.03	.30	-.06	-.02	-.64	-.58	-.28
Abnormal males	-.42	-.29	.11	-.26	-.19	-.60	-.60	-.37
Abnormal females	-.17	-.16	.17	-.21	-.13	-.63	-.58	-.38

Considering the relative unreliability of some of the MMPI variables, the intercorrelations of the K scale with other variables considered loaded with the K factor, are rather impressive: the G scale and plus scale were derived wholly by internal item relationships and without regard to criteria of any non-test behavior; the N scale corrected for self-criticality of certain plus-getters who showed deviant profiles; the Ch scale differentiated hypochondriacs from non-hypochondriacal abnormals who had elevated H scores; and a sub-set of items (Hy-O) which were chosen because they differentiated a clinical group--hysteria. There was, however, a considerable item overlap among these scales, which tended to raise the correlations. On the other hand, Meehl and Hathaway point out that the K scale was not actually pure for the hypothetical test-taking attitude because it was a composite of the test-taking scale L6 plus eight psychotic items. This would presumably tend to lower the correlations. Accordingly, Meehl and Hathaway substituted L6 for K, removed the item overlap among the scales C, N, Ch, L6 and Hy-O and calculated correlations among these reduced keys. Table III shows the intercorrelations among these five non-overlapping keys, based upon the responses of 150 unselected normal males between the ages of 26-45, rejecting records of F greater than 80. All scales were scored so as to render the correlations positive.[33]

---

[32] *Ibid.*, pp. 548-549.

---

[33] Ibid., pp. 551-552.

Table II

Correlations of K Scale with Other Variables Thought to Be Loaded with the K-Factor

	"+"	G	N	Ch	Hy-O
Normal males	-.64	-.76	-.70	-.67	.81
Normal females	-.62	-.73	-.64	-.63	.78
Abnormal males	-.70	-.75	-.69	-.64	.74
Abnormal females	-.70	-.81	-.72	-.71	.74

Table III

Intercorrelations of Five Scales Thought to Be Loaded with The Test-Taking Attitude, No Item Overlap, n=150 Normal Males

	G	Ch	L6	N
CH	.82			
L6	.76	.71		
N	.78	.73	.66	
Hy-O	.70	.63	.70	.59

This correlation matrix was subjected to a factor analysis and repeated three times in successively approximating the communalities because of the small number of tests. It appeared that one common factor was quite sufficient to account for the intercorrelations of these scales. The factor loadings of the scales

G, Ch, L6, N and Hy-O were .927, .868, .847, .818 and .770 respectively.[34]

It is claimed by Hathaway that the correction of the five clinical scales of the MMPI by use of K increases the proportion of clinically diagnosed cases scoring above the 90th percentile of normals.[35] Jeffery undertook an investigation of the factors which influence responses to K. She found that the differences characterizing high and low K were consistent for both the test and an interview situation. It was concluded that this observed set was persistent and deep-rooted. She states that the set evidences itself in items containing temporal adverbs. People who answered "true" tended to select categories containing the term "frequent" and the opposite appeared for those answering "false." Jeffery's data support the assumption that K measures two extremes of a deeply-rooted attitude causing distortion of personality items, but offered no explanation of the dynamics involved.[36]

McKinley, Hathaway and Meehl describe the statistical derivation of K weights as an optimal value which operates as a differential ratio in distinguishing a criterion group of abnormals diagnosed for each scale from a sample of unselected normals. The investigators emphasize, however, that this differentiation is only for the general population and state "it seems likely that for the best separation of maladjusted normals such as those which abound in a college counseling bureau and would be formally diagnosed in a psychiatric clinic as simple adult maladjustment, other weights might be better."[37]

---

[34] Ibid., pp. 552-553.

[35] S. R. Hathaway, Supplementary Manual for the MMPI (New York: Psychological Corporation, 1946).

[36] M. E. Jeffery, "Some Factors Influencing Answers on the Multiphasic Scale," (unpublished Doctor's dissertation, The University of Minnesota, 1946).

A study of the K scale by Schmidt using 98 Army convalescents who were diagnosed normal found that the L scale was as good an indicator of falsifying as K. He reports that the basic shape of the profile remained the same, but its height decreases with falsification. He concluded that the K factor contributes little, if anything, to differential diagnosis.[38] Hunt and others tried to check the discriminatory power of the K scale using psychotic and non-psychotic adult male psychiatric patients. They state that the K scale did not improve diagnosis and failed to reduce "false negatives."[39] In another study using 109 psychology students and 74 Navy general court martial prisoners Hunt found that the K correction failed to reduce deception. However, Hunt did find that an index of F minus K correctly identified malingerers, but not faked normal profiles.[40] Cofer and others found K and L scores significantly higher for faked normal profiles and suggested the use of an additive combination of L and K to identify these.[41]

---

[37] J. C. McKinley, S. R. Hathaway and P. E. Meehl, "The MMPI: VI. The K Scale," Journal of Consulting Psychology, 1948, 12:20-31.

[38] H. O. Schmidt, "Notes on the MMPI: The K Factor," Journal of Consulting Psychology, 1948, 12:337-342.

[39] H. F. Hunt, *et al.*, "A Study of the Differential Diagnostic Efficiency of the MMPI," Journal of Consulting Psychology, 1948, 12:331-336.

After reviewing MMPI literature Cottle concludes that K does not appear useful by itself to increase the discrimination of the clinical scales; however, the K scale in combination with L or F is useful to detect deliberate faking on the MMPI. In conclusion, Cottle states that there is evidence that the K scale reflects a persistent, deep-rooted attitude of distorting personality items and may be useful as a clinical scale to identify the defensive or the overly self-critical subject.[42]

Because of the contradictory evidence presented in the literature concerning the K scale and its function, no definite conclusion can be reached regarding its efficacy. Meehl and

---

[40] H. F. Hunt, "The Effect of Deliberate Deception on MMPI Performance," Journal of Consulting Psychology, 1948, 12:396-402.

[41] C. N. Cofer, J. Chance and A. J. Judson, "A Study of Malingering on the MMPI," Journal of Psychology, 1949, 27:491-499.

[42] W. C. Cottle, "The MMPI: A Review," Kansas Studies in Education, 1953, 3:9.

Hathaway's investigation of the K scale presented striking evidence of the validity of K to differentiate between deviant curves of actual normals and abnormals. However, in replication by other investigators the usefulness of K is not completely supported. All that can be said presently is that perhaps K or a fraction of K coupled with other validity key components can be useful in identifying certain test-taking attitudes.

Subtle-Obvious Keys (S-O)

While the K scale was being considered and developed Wiener was developing the concept of relatively subtle and obvious keys for the MMPI scales on the same ground as those given by Meehl and Hathaway above.[43] It was considered that the development of such keys on individual scales of the MMPI would yield more information and be of more practical usefulness than an overall validity scale, such as L, G and finally K.

Wiener considered that extremely deviate individuals could be picked out by a test consisting of obvious items. However, to help the counselor working with a normal population, it was thought that a much more subtle test was required which would both distinguish the extreme deviates and differentiate among the characteristics of a normal population. To Wiener these two services of a personality test appeared to be best served by developing both S and O scales.[44]

---

[43] D. N. Wiener, "Subtle and Obvious Keys for the MMPI," Journal of Consulting Psychology, 1948, 12:164-170.

In the development of the S and O keys the MMPI was used because 1) it was available for a relatively large normal population, 2) it was felt to have the most extensive and useful validation of any personality test, 3) it used generally accepted categories of personality characteristics, and 4) it had the unique feature of validity measures designed to indicate test-taking attitudes. It was felt, however, that when the MMPI was used with a relatively normal population, one that was functioning relatively successfully in society, certain important aspects of personality were masked because of its validation in terms of abnormal groups. The most obvious items distinguished the abnormal groups from the normals, whereas it appeared that the more subtle items should have had the greater validity in distinguishing the personality characteristics of normal groups.[45]

In developing the S and O keys all items of the MMPI were divided into two groups, those to which significant responses were relatively easy to detect as indicating emotional disturbance, and those to which they were relatively difficult to detect. The items for each scale were sorted into these two categories.

---

[44] *Ibid*., p. 164.

[45] *Ibid*.

All F scale items that also appeared in other scales were automatically assigned to the obvious category because by definition they seldom occur in a normal population. In addition, those items for which a blank response (no check on the answer sheet) was scored in a significant direction, were assigned to the subtle keys. Pooled judgments of raters was then used to sort the other items into the two categories with no attempt made to equalize the number of items in each group. There were more obvious items than subtle. The keys thus developed were used to rescore the test sheets of a representative sampling of 100 cases of the original male norm group for the MMPI. T-scores were developed and assigned to subtle and obvious item counts on the same basis as for the total scale T-scores. Subtlety-obviousness was determined rationally and not empirically; hence S-O scales were not formed for all MMPI subtests because Wiener felt that they contained too few S items.[46]

Tabulation tables for the raw scores on the subtle and obvious keys indicated a positive skew for most of the O item distributions of the norm group; relatively few individuals in the normal population answered the obvious items in a significant direction. It seemed probable to Wiener that significant answers to these O items were most characteristic of an institutionalized population. The S items were distributed in a relatively normal manner.

---

[46] Ibid., pp. 165-166.

An additional check on the validity of the selection of items for the keys was the frequency of their occurrence among the responses of a normal population. For all five scales the S items were answered in a significant direction approximately twice or more as frequently as the O items. The bases used to select items for the S and O keys for five scales of the MMPI yielded O items which were answered relatively frequently in a significant direction by a normal population compared with S items, and S items whose significant answers were in a reverse direction from the expectation of both the original authors of the MMPI.[47]

Recent research with the subtle and obvious scales for the MMPI raises some important interpretive problems.[48-51] While it was not puzzling to find that hospitalized psychiatric patients and other maladjusted and unsuccessful groups obtained high T-scores on the obvious scales, it was disconcerting to Fricke to find that groups of recovered psychiatric patients, successful trainees, successful salesmen and college sophomores obtained higher T-scores on the S scales than the normal MMPI population. In addition, these groups obtained higher overall T-scores than groups of unrecovered psychiatric patients, unsuccessful trainees, etc.[52] Because each of the items in the subtle scales was originally selected by Hathaway and McKinley due to its discrimination between normal and abnormal groups, it appears that there is something common either to the items or to the groups which influences the size of the T-scores more than does the original discriminating power of the items.

---

[47] Ibid., pp. 166-168.

[48] E. Rosen, "Self Appraisal, Personal Desirability, and Perceived Social Desirability of Personality Traits," Journal of Abnormal and Social Psychology, 1956, 52:151-158.

[49] E. Rosen, "Self Appraisal and Perceived Desirability of MMPI Personality Traits," Journal of Counseling Psychology, 1956, 3:44-51.

[50] D. N. Wiener, "Selecting Salesmen with Subtle-Obvious Keys for the MMPI," American Psychologist, 1948, 3:364.

[51] D. N. Wiener, "A Control Factor in Social Adjustment," Journal of Abnormal and Social Psychology, 1951, 46:3-8.

Fricke suggests that it is quite possible that groups obtaining high T-scores on the S scales have a response set to answer false. Fricke points out that Wiener's division of the MMPI items for five clinical scales into S and O items tends to make false the scored response for the S items. Four of the five scales have a majority of items scored false. Only the hypomania scale has more true than false responses and it is of interest to note that it is this scale which does not separate the groups studied by Wiener.[53]

---

[52] B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.

On the basis of some data and theoretical considerations, Wiener suggests that the S items are best for assessing the personality of normal persons and that the O items are best for abnormal persons.[54]

Fricke states:

> If it is true that S items are more likely to function in a normal population and O items in an abnormal population, then according to the present contention that K operates as a measure of response set, a difference in correlations should be reflected in normal and abnormal groups. Specifically it would be expected that the correlations of K with the five clinical scales utilized would be more positive or less negative in a normal than in an abnormal population. Correlations reveal that four of the five expectations are fulfilled; only the correlation between K and hypomania are not substantially different. A tentative hypothesis drawn from the data is that the more positive the correlation between scores from a measure of response set to answer false and scores from clinical scales composed of S-O items, the more likely it is that the group is well adjusted or successful.[55]

Set T Scale

According to Cronbach who has reviewed several response sets that influence a test-taker's behavior the variance generated by a response set regarded undesirable because it contributes only error variance and cannot be used to increase the usefulness of a test.[56-57] However, Fricke assumes a different rationale which stresses that response set can be used to improve the validity of many tests.[58]

---

[53] D. N. Wiener, "Subtle and Obvious Keys for the MMPI," Journal of Consulting Psychology, 1948, 12:168-169.

[54] Ibid., p. 170.

[55] B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:220-252.

In surveying the responses scored in most personality scales Fricke discovered that the significant responses are predominantly in one direction. For example, 96 out of the 100 items in the Cornell Index are keyed yes and 47 of 60 items in the hysteria scales of the MMPI are keyed false. A review of personality tests validated to predict academic achievement disclosed that for the more valid tests the response predictive of high achievement is usually false, no, or disagree. Some other tests in which a majority of scored responses fall into a particular response category are the California Psychological Inventory, the Humm-Wadsworth Temperament Analysis, the Minnesota Teacher Attitude Inventory, and the Strong Vocational Interest Blank.[59]

---

[56] L. J. Cronbach, "Response Sets and Test Validity," Educational and Psychological Measurement, 1946, 6:475-494.

[57] L. J. Cronbach, "Further Evidence on Response Sets and Tests Design," Educational and Psychological Measurement, 1950, 10:3-31.

[58] B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.

[59] B. G. Fricke, "The Development of an Empirically Validated Personality Test Employing Configural Analysis for the Prediction of Academic Achievement," (unpublished Doctor's dissertation, University of Minnesota, 1954).

Fricke contends that when the number of score responses is unequally divided between all possible response categories the effect of a response set may be substantial. If response set per se is not directly related to the criterion, it operates to introduce error in the criterion predictor. Fricke states that this is usually the case, but if response set is related to another variable that is criterion-related then response set may be used as a suppressor variable.[60] By suppressing or removing the influence of response set from the criterion-predictor an improvement in test validity can be effected.[61-64]

Fricke originally held the view that the imbalance of true and false items was simply a function of how the statements were written. But this seemed unlikely since on some tests having scales for several personality dimensions, the scored direction is different for different scales. It was concluded that the nature of the criterion groups and not the items was responsible for the imbalance. As a result Fricke set out to determine whether or not a measure of response set could be used to increase test validity.

---

[60] Ibid.

[61] Paul Horst, "The Prediction of Personal Adjustment," Social Science Research Council Bulletin, 1941, Number 48.

[62] A. S. Levine, "A Technique for Developing Suppression Tests," Educational and Psychological Measurement, 1952, 12: 313-315.

[63] Quinn McNemar, "The Mode of Operation of Suppressant Variables," American Journal of Psychology, 1945, 58:554-555.

[64] R. J. Wherry, "Test Selection and Suppressor Variables," Psychometrika, 1946, 11:239-247.

One index or measure of set for answering true was obtained by counting the times a person marked true to statements in a test. This method was utilized by Humm and Wadsworth to get a measure of suggestibility or cooperativeness. The counted the times a test-taker answered no to questions in the Temperament Schedule. Answer sheets having a large or small "no count" were not considered sufficiently valid for further analysis. According to their suggested cut-off points, 30-60 percent of the answer sheets were rejected as invalid.[65]

Fricke considered that a much more sensitive index of response set would be a count of the true responses for those statements which 49-51 percent of the test-takers marked true. For such statements there was no general agreement on an answer; they held maximum controversiality. A person without a response set would respond a true as often as false on these items. A person with a high score could be thought to have a strong set for marking true. A low score would indicate the opposite. Fricke emphasized that items for a set scale would be selected independently of any criterion external to the test. Because it was difficult to obtain enough items to form a scale at the 49-51 percent level of controversiality, it was decided to accept for the true response set scale (Set T) those statements in the Opinion, Attitude and Interest Survey which 40-60 percent of each criterion group marked true.[66] Hence, the Set T scale of the OAIS consisted of 69 statements scored in the true direction.

---

[65] D. G. Humm and G. W. Wadsworth, "The Humm-Wadsworth Temperament Schedule," American Journal of Psychiatry, 1935, 92:163-200.

Upon completion of his investigation, Fricke concluded that a measure of response set could be used to increase greatly the accuracy of criterion prediction.[67]

The B Scale

The B scale of the MMPI is structurally and functionally similar to the K scale but differs markedly from K in the method used for its construction. B, a measure of response bias, was modeled after the Set T scale of the OAIS.[68] The B scale and the Set T scale made it possible to measure a test-taker's tendency to answer true to statements in a personality inventory. Individuals with a strong bias to answer true obtain low hysteria scores on the MMPI due to the fact that 78 percent of the Hy items are scored false.[69] Fricke felt that if a fraction of a Set T type of scale was added to Hy or if a fraction of K was subtracted from Hy, the validity of Hy would be improved. The assumption was that by partialing out, or suppressing the influence of response bias, the purity of the clinical scale would be increased.[70]

---

[66] B. G. Fricke, The Opinion, Attitude and Interest Survey (Minneapolis: Investors Diversified Services, 1955).

[67] B. G. Fricke, "Response Set as A Suppressor Variable in the OAIS and MMPI," Journal of Consulting Psychology, 1956, 20:161-169.

[68] B. G. Fricke, The Opinion, Attitude and Interest Survey (Minneapolis: Investors Diversified Services, 1955).

To select items which would reflect response bias on the MMPI Fricke examined the item responses of normal persons. His objective was to obtain a pool of items of high controversiality, that is, items to which about equal numbers answer true and false. Items drawing true answers from 40-60 percent were considered sufficiently sensitive to be useful. Since some normals did not answer true or false to every item, half the "cannot say" items were added to the true answers to establish whether or not each item met the arbitrary 40-60 percent level of controversiality. Fricke considered that a test-taker with a strong bias to answer true would be expected to achieve a high score when all the controversial items were scored.

---

[69] B. G. Fricke, "Conversion Hysterics and the MMPI," _Journal of Clinical Psychology_, 1956, 12:322-326.

[70] B. G. Fricke, "A Response Bias Scale for the MMPI," _Journal of Counseling Psychology_, 1957, 4:149-153.

Two normal samples were involved. The first sample of 604 cases was used by Hathaway and McKinley in the construction of the clinical scales; it consisted of a sub-group of 339 Minnesota normals and a sub-group of 265 college normals. The percentage of these two sub-groups who answered true to each item was averaged. The second sample of 589 cases was used as the norm group for the more recently constructed non-clinical scales; it consisted of a sub-group of 253 normal males and a sub-group of 336 normals females. The percentages of these two sub-groups also were averaged. A total of 81 items were located which were answered true by 40-60 percent of both normal samples. Of the 81 items of high controversiality 18 were found to be in the 30-item K scale. The response bias scale B consists of the 63 non-K items. Fricke states that high T-scores on B indicate a tendency to answer false.

Because of the complete lack of evidence in the literature concerning Fricke's interpretation of the B scale, no definite conclusion can be reached regarding the efficacy of the B scale.

Miscellaneous Scales

The G scale. About three years before research on the test-taking attitude was begun, Hathaway and Estes, using a variant of the method of internal consistency, developed a scale called G. This scale was the only MMPI scale which was derived without the use of a criterion external to the test; the selection and scoring of items was based wholly upon the intercorrelations among the items themselves. Essentially, the procedure consisted in locating among a group of 101 unselected normals those individuals who, when their answer sheets were used as scoring keys, produced the maximum variance of the other 100 scores. The assumption was that these persons were the most extreme deviates on whatever factor or factors contributed most heavily to the variance and covariance of the total pool of MMPI items. From the evidence adduced by Mosier, it is of course clear that the purity of factorial unity of this hypothetical underlying continuum is by no means guaranteed by such a procedure.[71] Meehl and Hathaway state that this maximizes the variance of a set of items by scoring them in such a direction as to maximize their mean covariance (since the item variances are unaffected by the direction of scoring).[72] Instead of actually calculating the variances for the 2^{550} ways of scoring the test, the investigators selected individuals who approximated the optimal scoring key. It was found that the scoring keys for some ten individuals selected by this method tended to form two distinct clusters, each of which consisted of keys (individuals) showing high correlations with one another and high negative correlations with the members of the other cluster. An item analysis was then carried out on these two small groups, and the items resulting were combined into a scale called G (general factor).

---

[71] C. I. Mosier, "A Note on Item Analysis and the Criterion of Internal Consistency," Psychometrika, 1936, 1:275-282.

[72] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:549-550.

The G scale, although derived without recourse to any clinical group whatever, nevertheless showed a correlation of .91 with clinical scale Pt. The mean MMPI curves for unselected normals with high G (the neurotic end) showed elevations on seven MMPI clinical scales and on F; whereas L (raw score) and Hy tended to fall below the mean. The mean profile for normals with low G was almost an exact mirror image of this curve. However, G was not found to be effective in the detection of any clinical group or to be particularly useful for any purpose; and because at that time no theoretical basis was available for interpreting it, the scale was abandoned. Another scale, called plus was derived in a similar but not identical manner.[73]

The N scale. When Meehl and Hathaway first discovered that certain clearly abnormal persons obtained normal MMPI profiles, and on the other hand, that certain normals obtained elevated profiles with no evidence of deliberate falsification, they began an investigation which initially led to the development of the N scale.

---

[73] *Ibid.*, p. 550.

The key discriminated individuals who were overly critical in reporting themselves (plus-getters) from actually abnormal subjects with similar MMPI profiles. This scale, however, did not prove to be useful in detecting impunitive or defensive sortings. For this reason, further study was carried out which resulted ultimately in the development of K.[74]

The Ch scale. In the derivation of the original hypochondriasis key, there was developed a correction scale called Ch, the function of which was to separate actual clinical hypochondriacs from a group of non-hypochondriacal abnormals who attained spuriously elevated scores on H. The item content of this Ch key was quite puzzling to Hathaway and McKinley because, although the correction was successful, the items did not appear to refer to anything either hypochondriacal or anti-hypochondriacal. They possessed no apparent psychological homogeneity. The majority of the items on Ch were scored if answered in the statistically rare and obviously maladjusted direction. They apparently measured some non-somatic component of test responses which resulted in spuriously elevated H scores in persons who were not actually hypochondriacal.[75]

---

[74] S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule: I. Construction of the Schedule," Journal of Psychology, 1940, 10:249-254.

[75] P. E. Meehl and S. R. Hathaway, "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:550-551.

The Hy-O scale. The Hy-O items reflected the N component, although they were scored in the opposite direction from N, as well as from Ch and G. Hy-O consist of a sub-set of items which were chosen because they differentiated empirically the hysteria clinical group from normals. It was concluded that these items reflected the self-deceptive and impunitive attitude of the hysterical temperament.

The Cd and the Hy-S keys. Other minor attempts of validity key construction on the MMPI consists of the Cd and Hy-S scales. The former was developed to differentiate systematic depressives from normals; the latter items constituted an attempt to distinguish hysteria, hypomania and psychopathic deviates from normals.[76,77]

Consistency keys. Consistency or verification of response checks are no longer infrequent keys in various objective personality and interest inventories. The instruments devised by Kuder, Strong, and Edwards typify contemporary attempts to uncover falsification by using consistency of response to determine whether the subject understood test directions, possessed

---

[76] S. R. Hathaway and J. C. McKinley, "A Multiphasic Personality Schedule: III The Measurement of Symptomatic Depression," Journal of Psychology, 1942, 14:73-84.

[77] J. C. McKinley and S. R. Hathaway, "A Multiphasic Personality Schedule: V Hysteria, Hypomania and Psychopathic Deviate," Journal of Applied Psychology, 1944, 28:153-174.

sufficient reading skills to comprehend the test items, or answered carelessly or insincerely.[78-80]

Similarity Between Various Scales

B and K scales. Although the methods used in construction of the B and K scales were vastly different, structurally and functionally B and K are quite similar. Both consist of items of high controversiality and both have their items scored in one direction.[81]

The functional similarity of B and K is revealed by the correlation coefficients shown in Table IV. The correlation between B and K for 336 normal females is minus .68 and the correlation for 253 normal males is minus .67. The correlation for a sample of 63 conversion hysterics was found to be minus .73.[82]

Fricke concluded from the data in Table IV that B and K were the two most similar validity indicators. It is of interest

---

[78] G. F. Kuder, Kuder Preference Record: Vocational (Chicago: Science Research Associates, 1956), p. 3.

[79] G. F. Kuder, Kuder Preference Record: Occupational (Chicago: Science Research Associates, 1959), p. 4.

[80] A. L. Edwards, Edwards Personal Preference Schedule (New York: The Psychological Corporation, 1959), p. 15.

[81] B. G. Fricke, "A Response Bias Scale for the MMPI," Journal of Counseling Psychology, 1957, 4:150.

[82] Ibid., p. 151.

Table IV

Intercorrelations of Four MMPI Validity Scale Indicators and Hy for Normals

	Females		Males		
	L	F	K	B	Hy
L		-.28	-.38	-.30	-.13
F	-.01		-.36	.40	.19
K	.36	-.35		-.68	.29
B	-.28	.43	-.67		-.23
HY	.19	.04	.39	-.37	

to note that the correlations between K and L are higher than the correlations between B and L and that the correlations between B and F are higher than the correlations between K and F. From this Fricke interpreted that the L component was stronger in B than in K.

The correlations between Hy and B, and between Hy and K suggest that B and K could be used as suppressor variables. Fricke suggests that the Hy scale could be improved by adding to it a certain fraction of B and by subtracting from it a certain fraction of K. However, what values should be used were not discussed.[83]

---

[83] B. G. Fricke, "Subtle and Obvious Test Items and Response Set," Journal of Consulting Psychology, 1957, 21:250-252.

B and K as discriminators. A function more influential than the suppressor action is evidently operating in B and K. Fricke states that B and K are both unsuccessful suppressors because they both discriminate conversion hysterics from normals; a low B and a high K are indicative of conversion hysteria. Consequently, this investigator concludes that the subtraction of a fraction of B from Hy and the addition of a fraction of K to Hy would improve the validity of Hy if B and K tap something diagnostic of conversion hysteria that is not tapped by Hy.[84]

It is important to note that B and K each appear to function simultaneously as scale suppressor and criterion discriminator. This is unfortunate since the suppressor and discriminator effects are in opposite directions and tend to cancel each other. The influence of B and K as suppressors is weaker than their influence as discriminators and this results finally in the subtraction of B and the addition of K.[85]

Because the level of K scores, and probably B, is affected by the social-educational-economic level of the test-takers, it was Fricke's speculation that for the lower levels the discriminator role of K (and B) would be much more important than the suppressor role, but that for the higher levels (college students) the discriminator role of K (and B) would be much less important than the suppressor role. A possible explanation for McKinley, Hathaway and Meehl's finding that the addition of a fraction of K to Hy did not improve Hy is to be found here. If their conversion hysterics and normal control cases had the same mean K scores, then the addition of a fraction of K probably would have capitalized on K's suppressor role and improved the validity of Hy.[86]

---

[84] B. G. Fricke, "A Response Bias Scale for the MMPI," Journal of Counseling Psychology, 1957, 4:152.

[85] Ibid.

Discrepancies in B and K scores may be of some importance (that is, it is possible that a test-taker with a T score of 70 on B and a T score of 60 on K is less defensive than a test-taker with a T score of 50 on B and a T score of 60 on K; it might be argued that the first test-taker's K score was obtained largely due to his response set to answer false but that the second test-taker's K score was obtained through the operation of something other than response set). The assumption is that K is more than a measure of response set. If the assumption is not valid, there probably is no need for the two test scores.[87]

While both scales appear to reflect to a certain extent a test-taker's nontest behavior, their primary role is to reveal something about a test-taker's test taking behavior so that more accurate inferences can be drawn from the diagnostic clinical scales.

---

[86] Ibid.

[87] Ibid.

Functional similarity of K and Set T. The functional similarity of K and Set T scales is seen in the fact that two-thirds of all K items fall in the controversial range. Further evidence on the similarity of Set T and K is found in their correlation with each other: minus .58 to minus .71. To Fricke, it appeared that Set T and K accomplished essentially the same thing. Whether Set T was superior to K was not determinable from the available data. While K may not be as pure a measure of response set, it may do something in addition to what is done by Set T.[88] Fricke concludes:

> A measure of response set such as K or Set T may function as a suppressor when there is no true-false imbalance, and may not function as a suppressor when there is a moderate true-false imbalance. Too frequently scores from the K scale have been correlated with scores from other tests and judgments have been made as to what K was "really" measuring. I don't know what K measures, no one else does either. It seems unlikely that response set keys would be able to bring validity to tests which are not validated empirically even though response set may be a major component in a test-taker's score.[89]

The marked structural and functional similarity of Set T of the OAIS and K of the MMPI was drawn upon to challenge the traditional interpretation of the K scale. Evidence was assembled which indicates that some of the MMPI scales are not optimally K corrected.

---

[88] Ibid.

[89] Ibid.

S and O relationships. Relationships exist between the S and O, and the L scale, intelligence, psychological sophistication and neuro-psychiatric diagnosis. The group with high scores on the L scale of the MMPI was higher on the S keys than was the low L scale group. For the group with low L scores, the O scores were for all scales approximately equal to or higher than the S scores. Individuals of high ability (intelligence T score above 60 on the Otis) have approximately equal O and S scores, whereas individuals of low ability (T score below 40) have generally higher O scores than S, and higher O scores than the high ability group.[90]

MMPI profiles of a psychologically sophisticated group showed distinction between S and O keys, with S much higher than O whether the group was giving honest results or was attempting to fake good. With this sophisticated group it appeared to make little difference whether the test was taken honestly or faked good; in either case, O items were successfully avoided whereas S items yielded average and above average T scores.[91]

In general, O keys are highly correlated with each other and have no correlation with S keys; S keys have a low positive correlation with each other. There is a high negative correlation between O minus S and K, indicating the considerable weighting of K with S items. There is evidence that high L scale scores are associated with higher S than O scores, whereas the converse is true for low L scores; psychologically sophisticated individuals almost completely avoid significant O responses and have much higher S scores. While total scale scores on the MMPI failed to differentiate significantly between successful and unsuccessful students and on the job trainees, O keys were significantly higher than the S for the unsuccessful group. The S keys were insignificantly lower, and the total scale scores were between the two.[92]

---

[90] D. N. Wiener, "Subtle and Obvious Keys for the MMPI," Journal of Consulting Psychology, 1948, 12:169.

[91] Ibid., p. 170.

Summary

All major validity scales have been examined in this chapter. The order of discussion of the various keys was based upon the chronology of their development with the earliest devised scales discussed first. In addition to the major scales, however, several minor scales had to be considered in order to bring about a more lucid understanding of the major keys. All of the scales were developed around the MMPI with the exception of Set T. Set T was constructed with OAIS data.

Essentially, 1) the L scale used items which were socially desirable but rarely true of an individual; 2) the F scale was developed with rarity items (items chosen by ten percent or fewer of the sample); 3) the K scale items were selected because they differentiated normal individuals who scored abnormally from abnormal subjects who scored normally; 4) the S-O keys distinguished abnormal subjects from normal by the use of a subtlety-obviousness continuum; and 5) Set T and B scale items were developed by uncovering high controversial items and measured response set.

---

[92] Ibid.

Clear probability statements concerning the validity of the various scales are lacking.

The motivational research study with its instrumentation and the procedures utilized in the present investigation are discussed in Chapter III.

CHAPTER III

PROCEDURES

The present investigation used the administered test protocols of participants in Farquhar's motivational research project.[1] In this chapter the background, theory, design and instrumentation of Farquhar's project are discussed and the procedures for the present study are outlined.

Background of Farquhar's Study

In an attempt to collate a thorough objectively validated description of the personal characteristics of high and low academically motivated students, Farquhar strutinized the existing literature on motivation. As a result, a theory of need-achievement and non-need-achievement motivation was formulated. A summary of the basic motivation theiry is found in Tables V and VI.

Table V

Summary of Theory of Need-Achievement and Non-Need-Achievement Motivation Basic to Current Research

Motivational Situation	
Need-Achievement	Non-Need-Achievement
1. Long term involvement	1. Short term involvement
2. Competition with a maximal standard of excellence	2. Competition with a minimal standard
3. Unique accomplishment	3. Common accomplishment

---

[1] William W. Farquhar, "Motivation Factors Related to Academic Achievement," (East Lansing: Cooperative Research Project No. 846, Office of Education, United States Department of Health, Education, and Welfare, January, 1963), 506 pp.

Table VI

Hypothesized Personality Factors Associated with Academic Achievement

Academic Anxiety
Self Value
Authority Relations
Independence-Dependence Conflict
Activity Patterns
Goal Orientation
Interpersonal Relationships

Three attitudinal areas were ascertained which were considered capable of differentiating between over-and under-achievers. These areas consisted of attitudes toward school and learning, toward self and toward parents. Later, another area was established: attitudes toward occupations.

In order to test the posited motivational theory 725 items were logically developed with the purpose of contributing significantly to the regression of an academic predictor on grade achievement. The essential assumption was that students who exemplified extremes in academic performance should also exhibit extreme response to motivational characteristics.

General Design of the Motivational Study

The above hypothesis was tested with a sample of approximately 4200 eleventh grade Michigan public high school students who were attending nine separate institutions. Under- and over-achieving students were identified by the following procedure:
1) Schools which had 9th grade Differential Aptitude Test scores

available on their current 10th graders were contacted and asked to cooperate in the study. 2) A second aptitude measure was obtained so that reliable estimates of academic aptitude could be made. California Tests of Mental Maturity were administered to schools lacking this data. 3) DAT-Verbal Reasoning subtest and the CTMM Language scores were used in obtaining a stable estimate of academic aptitude after empirically examining possible DAT CTMM sub-score combinations. 4) Regression lines were calculated for each school and sex assuming a perfect correlation between DAT-VR and CTMM-L sub-tests. Separate equations were calculated because a pilot study indicated that one equation could not be generalized from school to school. Only those individuals who fell within one standard error of estimate above and below the regression line were included in the study. The methodology of selection of individuals with stable measured aptitude is summarized in Figure I. Because it was important that the criterion groups be classified with little chance of making a Type I error (accepting instead of rejecting) it was decided to run the risk of Type II error (rejecting instead of accepting) even if sample were lost in the process. 5) Regression equations predicting grade point average from DAT-VR scores were calculated for each sex in each participating school. The DAT-VR was used because it was found to correlate consistently higher with grade-point average than the CTMM-L scores. Under-

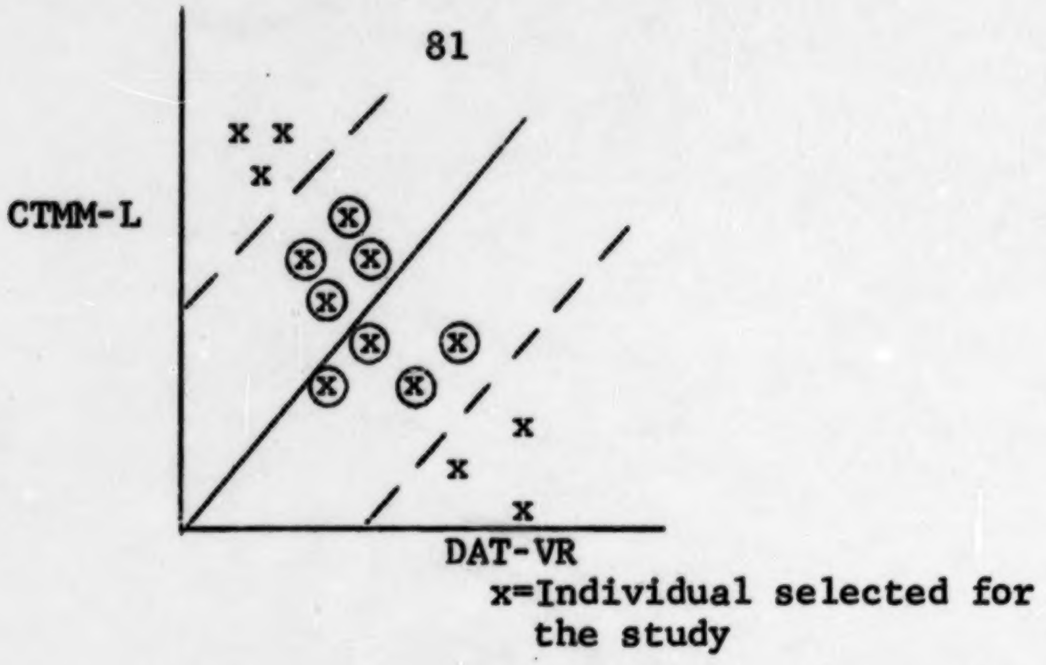
CTMM-L
DAT-VR
x=Individual selected for the study

Figure I. Methodological Selection of Individuals with Stable Measured Aptitude

achievers were defined as those individuals whose grade point average fell at least one standard error of estimate below the regression line prediction of achievement. Similarly, over-achievers were designated as falling one standard error above the regression line. The methodology used in selecting under- and over-achievers is summarizediin Figure II.

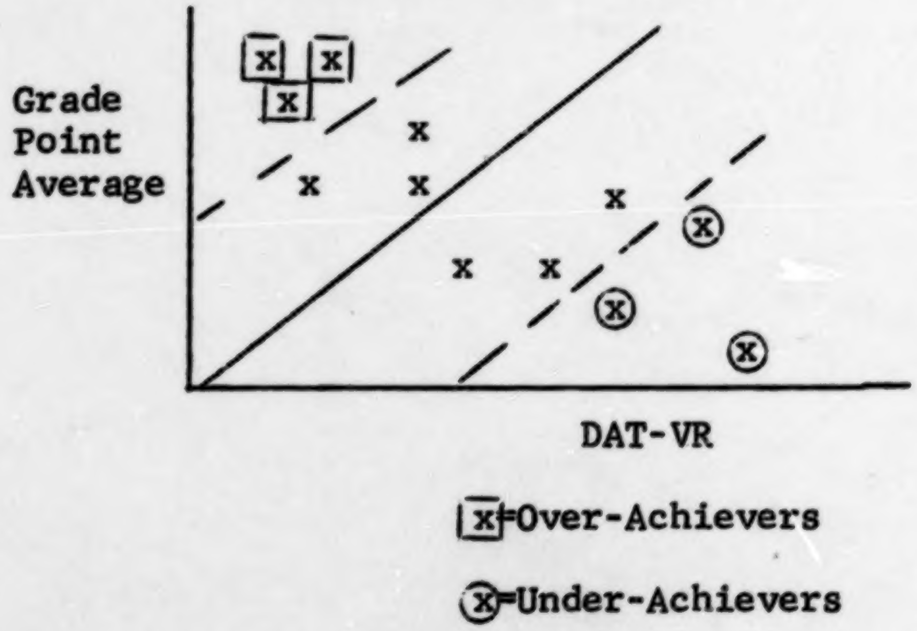
Grade Point Average
DAT-VR
x=Over-Achievers
x=Under-Achievers

Figure II. Method of Selecting Under- and Over-Achievers

By using the above method under- and over-achievers were selected from the full range of academic aptitude. Approximately twelve percent of the sample was classified in one of the extreme groups.

Instrumentation

The Generalized Situational Choice Inventory was developed by reviewing studies of over- and under-achieving students to determine characteristic differences. These characteristics were incorporated in pairs of statements which typified need or non-need-achievement. Classification into need or non-need-achievement was based on face validity of the content of choice. Choices had to meet the following criteria: 1) be as free as possible of culture stereotype; 2) allow the individual to project his choice into the future; and 3) be purely as possible need-achievement or non-need-achievement. It was not necessary that the choice be concerned with only one of the three sub-factors provided it could be clearly designated need or non-need achievement. The inventory consists of 200 item pairs.

The Preferred Job Characteristics Scale was developed essentially in the samemmanner as the inventory above. Previous studies indicate that occupational goals are positively correlated with need-achievement.[2] The scale consists of 64 items.

---

[2] *Ibid.*

The Human Trait Inventory was developed by reviewing items from personality measures which previous research found to differentiate between under- and over-achieving students. The inventory consists of 120 items.

The Word Rating List was developed by using adjectives which described under- and over-achievers. The inventory consists of 110 items.

Procedures for the Present Investigation

The following procedures were used in carrying out the present study:

1) M-Scale items which differentiated between over- and under-achievers were determined for all male and female over-achievers and for all male and female under-achievers in Farquhar's study.

2) An item frequency distribution was constructed for each group for each sex. The point of overlap where under-achievers scored as over-achievers on all battery discriminating items (items which differentiated under-achievers from over-achievers) was determined by plotting the respective normal distribution curves. The point of overlap which was to be identified is illustrated in Figure III.

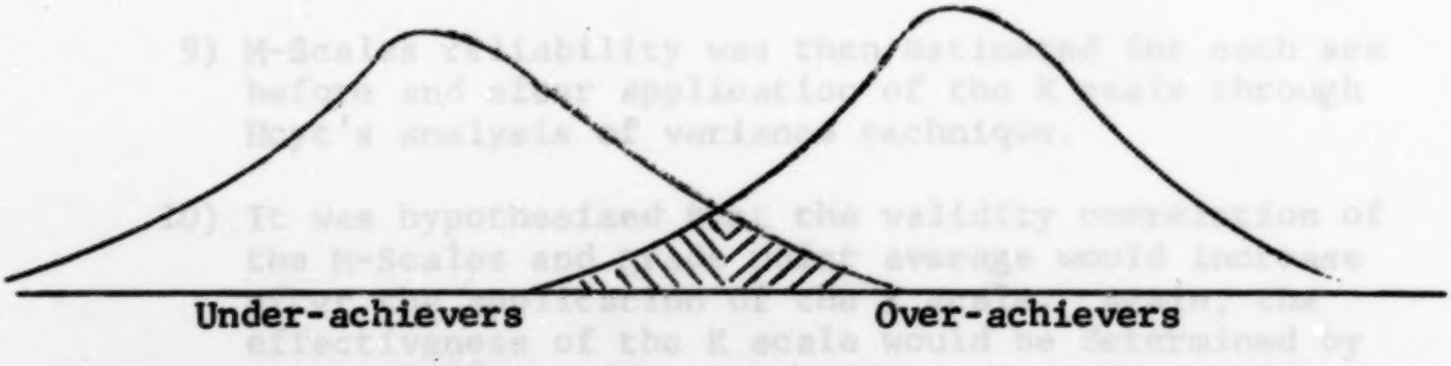

Figure III. Theorized Model of Selection of Misclassified Under-achievers and Over-achievers Differentiated by the M-Scales.

3) Under-achievers who possess M-Scales scores as large as or greater than the overlap point were identified. These under-achievers scored as over-achievers on M-Scales discriminating items. In addition, the over-achievers who selected less than the over-lap point scored as under-achievers on M-Scales discriminating items. The above groups for each sex were then labeled misclassified over-achievers and misclassified under-achievers, respectively.

4) The protocols of the misclassified over- and under-achievers were then analyzed by total battery by items by the Chi-square technique.

5) After item analysis M-Scales items which were significant at the .05 level of confidence or higher were selected for inclusion into the K scale for each sex.

6) Reliability of the K scale was then determined by Hoyt's analysis of variance technique. Stratified general population random samples of 100 males and 100 females, treated separately, were used in determining K item reliability. In addition, replicative samples of 100 males and 100 females, treated separately, were used in cross-validating reliability.

7) To obtain evidence of K scale validity two approaches were examined for the effect of K on 1) M-Scales reliability and 2) M-Scales validity.

8) It was hypothesized that the M-Scales reliability would increase after the application of the K scale because homogeneity of test performance would increase. The effectiveness of the K scale would be determined by its ability to remove defensive, threatened individuals who would distort and hence lower the test reliability coefficients.

9) M-Scales reliability was then estimated for each sex before and after application of the K scale through Hoyt's analysis of variance technique.

10) It was hypothesized that the validity correlation of the M-Scales and grade point average would increase after the application of the K scale. Again, the effectiveness of the K scale would be determined by its ability to remove test-defensive individuals.

11) M-Scales validity was then estimated for each sex before and after application of the K scale through the Pearsonian correlational technique.

In this chapter, the background, theory, design, and instrumentation of Farquhar's motivational project were discussed. In addition the design of the present investigation was outlined. The results of the study are discussed in Chapter IV.

CHAPTER IV

RESULTS OF THE INVESTIGATION

The outcomes of the present investigation are discussed in this chapter. The selection of the items comprising the K scale, sex differences in K item selection, reliability and validation of the scale are presented.

The number of K items in each test with sex differences in selection is illustrated in Table VII.

Table VII

Number of K Items in Each Test with Sex Differences in Selection

GSCI		Word Rating List	
Males	8	Males	20
Females	16	Females	32
Human Trait		Preferred Job	
Males	7	Males	7
Females	13	Females	26

Distribution of K Items

A K item frequency distribution was constructed for a stratified general population sample of 100 males and 100 females with a similar cross-validation sample of 100 males and 100 females. Of a total of 42 possible K choices only three percent of the males selected no K item. Sixty-five percent of the sample chose fifteen or less K responses and eighty-two percent selected twenty or less items. The highest number of K items selected by any one individual was thirty. The male K item frequency distribution is illustrated in Figure IV.

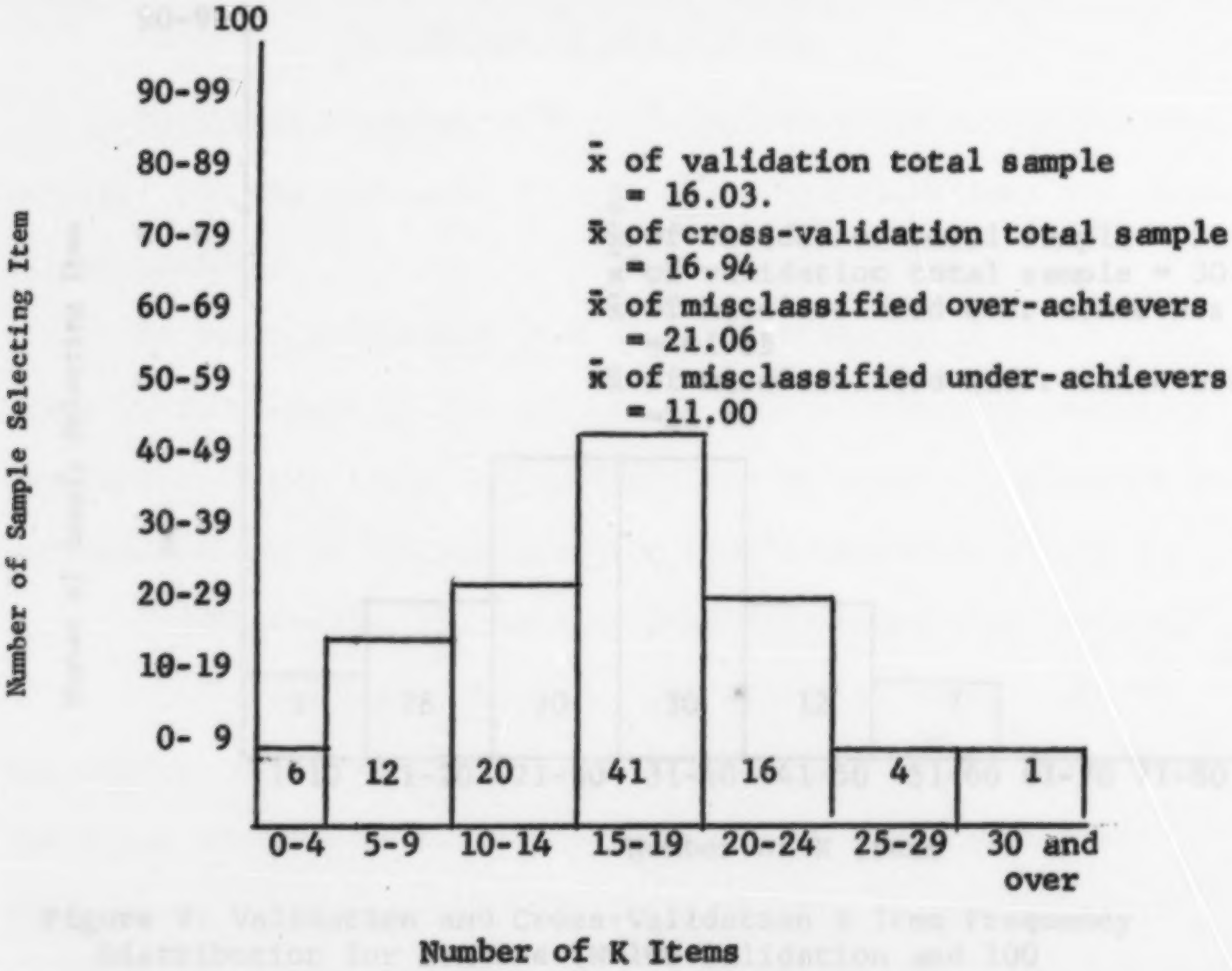

Figure IV. Validation and Cross-Validation K Item Frequency Distribution for Males (N=100 validation and 100 cross-validation)

Of a total of 87 possible K choices for the females all females selected no fewer than ten K items. Eighty-one percent of the sample chose forty or less K items and ninety-three percent selected fifty or fewer items. The highest number of K items selected by any one individual was fifty-seven. The female K item frequency distribution is illustrated in Figure V.

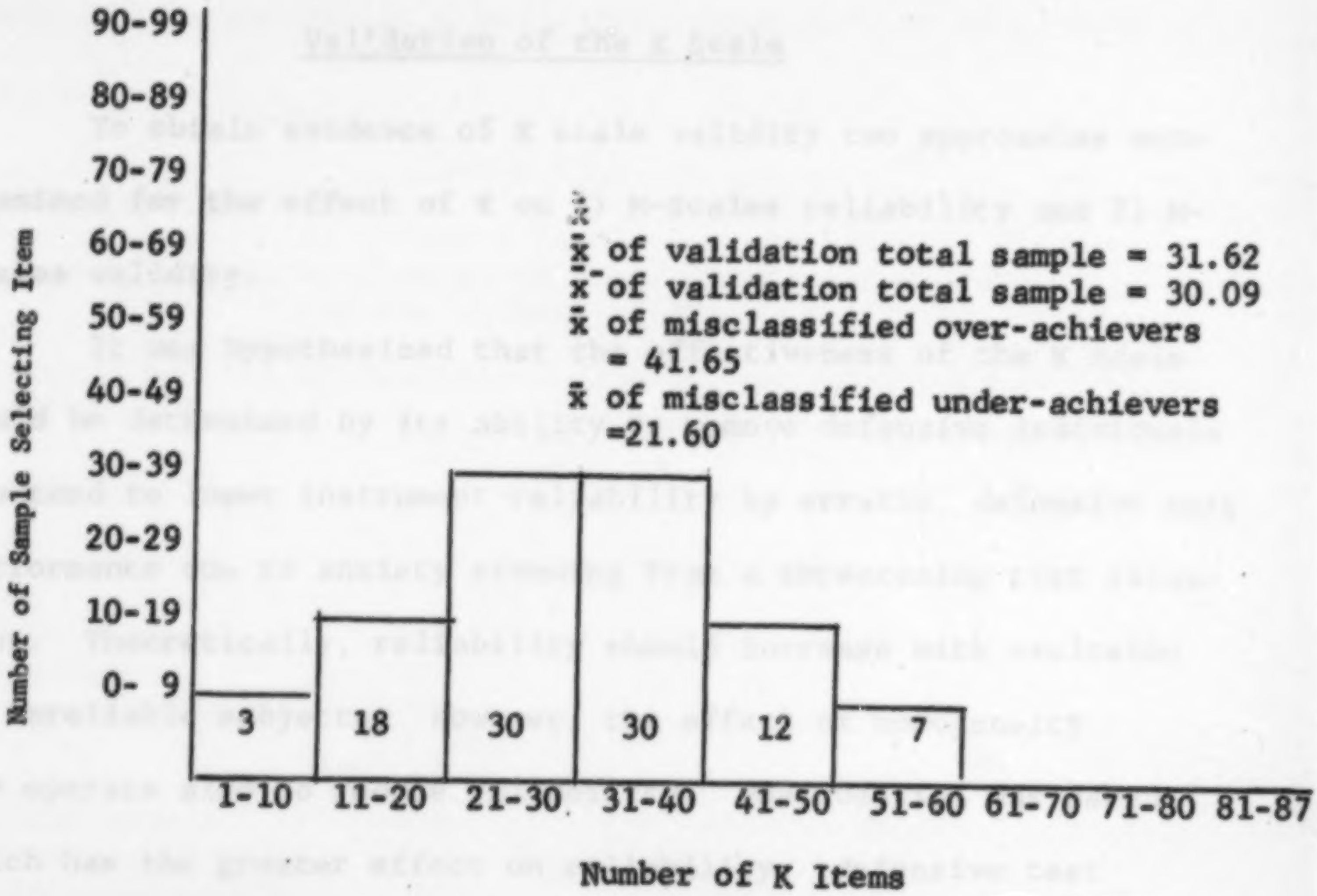

Figure V. Validation and Cross-Validation K Item Frequency Distribution for Females (N=100 validation and 100 cross-validation)

Reliability of the K Scale

Hoyt's analysis of variance technique was used in estimating K scale reliability. Reliability was based on a stratified random sample of 100 males and 100 females. A cross-validation reliability check was also carried out using again 100 males and 100 females in a stratified random sample. For the validation and cross-validation male groups, reliability estimates of .734 and .799 were obtained, respectively. For the females, reliability estimates of .884 and .856 were determined, respectively.

Validation of the K Scale

To obtain evidence of K scale validity two approaches were examined for the effect of K on 1) M-Scales reliability and 2) M-Scales validity.

It was hypothesized that the effectiveness of the K scale could be determined by its ability to remove defensive individuals who tend to lower instrument reliability by erratic, defensive test performance due to anxiety stemming from a threatening test situation. Theoretically, reliability should increase with exclusion of unreliable subjects. However, the effect of homogeneity may operate also to reduce reliability. The question was asked which has the greater effect on reliability: defensive test performance or homogeneity? To test the effects of both of the above statements a stratified random sample of 100 males and 100 females was selected. A cross-validation group reflecting the above numbers was also used. Individuals possessing K scores as large as or greater than a T-score of 65 (for the males this represents a raw score of twenty-three and for the females a raw score of forty-seven) were excluded from the sample. In addition, an equal sized sample was randomly excluded to determine its effect on homogeneity. Hoyt's analysis of variance technique for estimating internal consistency reliability was used to obtain an estimate of consistency of the M-Scales before and after application of the K scale. The effects on reliability after application of the K scale are summarized in Tables VIII and IX.

Table VIII

Effects on M-Scales Internal Consistency Reliability Before and After Application of the Male and Female K Scale for the Validation Group

	Before Application of K		After Application of K	
	n	r	n	r
Males				
Total Sample	100	.842	95	.873
Randomly Excluded Sample	95	.808		
Females				
Total Sample	100	.837	88	.926
Randomly Excluded Sample	88	.785		

Table IX

Effects on M-Scales Internal Consistency Reliability Before and After Application of the Male and Female K Scale for the Cross-Validation Group

	Before Application of K		After Application of K	
	n	r	n	r
Males				
Total Sample	100	.887	91	.911

Table IX (continued)

	Before Application of K		After Application of K	
	n	r	n	r
Randomly Excluded Sample	91	.823		
Females				
Total Sample	100	.894	84	.937
Randomly Excluded Sample	84	.876		

Although no significant differences between reliability coefficients were found, the magnitudinal direction of the coefficients did substantiate the hypothesis: as previously stated, after application of the K scale instrument reliability should increase. In both male and female validation and cross-validation samples the magnitude of reliability coefficient increased after application of K.

Effect of K on Validity

The effect of the K scale on the validity coefficient between M-Scales raw scores and standardized grade point averages was also determined. It was hypothesized that the Pearsonian validity coefficient between the two variables would increase after application of K. The effects of K on validity for males and females are summarized in Table X.

Table X

Effects on the Validity Coefficient Between M-Scales Raw Scores and Standardized Grade Point Averages After Application of the Male and Female K Scale for Validation and Cross-Validation Groups

	Validation Group		Cross-Validation Group	
	n	r	n	r
Males				
Correlation Before Application of K	100	.573	100	.493
Correlation After Application of K	95	.610	91	.531
Correlation After Random Exclusion	95	.511	91	.458
Females				
Correlation Before Application of K	100	.487	100	.574
Correlation After Application of K	88	.556	84	.611
Correlation After Random Exclusion	88	.513	84	.532

No significant differences between validity coefficients were found after the application of K. However, in both male and female validation and cross-validation samples validity coefficient magnitude increased after K was applied. All validity correlations were significant from zero at the 3 percent or better level of confidence.

Summary

The outcomes of the present investigation were presented in this chapter. Selection of the M-Scales items comprising the final K scale was based upon an item analysis of the total battery items of the misclassified over-achievers and under-achievers by sex. M-Scales items which were answered differently by each of the above groups at a significance level of .05 or better were selected for inclusion into the K scale. In the final form the male K scale consists of 42 items. The female K scale is composed of 87 K items.

Frequency distributions for each sex were determined. The range of K item selection for the male sample was from zero to 30. The mean number of K items for the males was determined to be 16. For the females K item selection ranged from a low of 10 to a high of 57 with the mean falling at 32.

Hoyt's analysis of variance technique was used in determining K scale reliability. For the males reliability estimates of .734 and .799 were obtained on validation and cross-validation samples, respectively. For the females reliability coefficients of .884 and .856 were determined respectively for validation and cross-validation samples.

To obtain evidence of K scale validity two approaches were examined for the effect of K on 1) test reliability and 2) test validity.

Individuals possessing K item T-scores of 65 or greater were excluded from the sample. Hoyt's analysis of variance technique for estimating internal consistency was used to obtain a reliability statement of the M-Scales before and after application of the K scale. For both male and female validation and cross-validation samples no significant differences in reliability coefficients were obtained, although in all cases the coefficient increased in magnitude after the application of K.

The effects on validity correlation of M-Scales raw scores and standardized grade point averages before and after application of the male and female K scale were determined. Before application of K the male validation and cross-validation validity coefficients were .573 and .493, respectively. After use of the K scale the male correlations increased to .610 and .531 for the validation and cross-validation groups, respectively. For the females validation and cross-validation coefficients of .487 and .574 were obtained before use of K. After application of the K scale these correlations increased to .556 and .611, respectively. However, no significant differences in correlations were obtained after K was applied. All validity correlations were significant from zero at the 3 percent or better level of confidence.

The concept of the K scale was corroborated, however, from the standpoint of the rationale of defensive test-taking attitudes. It was expected that the misclassified over-achiever who in essence is an under-achiever would be more threatened by the test situation and hence, would be more defensive. With such an attitudinal overlay this individual would choose more K items. This expectancy was significantly found to be the case for both male and female groups.

The interpretation of the findings and the summary of the investigation are presented in Chapter V.

CHAPTER V

SUMMARY AND CONCLUSIONS

The Problem

This investigation was concerned with the development and validation of a K scale for an objective test battery on motivation.

Methodology

The present investigation used the administered test protocols of participants in Farquhar's motivational research project. Approximately 4200 eleventh grade Michigan public high school students comprised the population from which sample was drawn for this study.

Instrumentation consisted of the Generalized Situational Choice Inventory, the Preferred Job Characteristics Scale, the Human Trait Inventory and the Word Rating List. The battery consists of 502 items.

Selection of the M-Scales items comprising the final K scale was based upon an item analysis of the total battery items of the misclassified over-achievers and misclassified under-achievers by sex. M-Scales items which were answered differently by each of the above groups at a significance level of .05 or better were selected for inclusion into the K scale. In the final form the male K scale consists of 42 items. The female K scale is composed of 87 K items.

Frequency distributions for each sex were determined. The range of K item selection for the male sample was from zero to 30. The mean number of K items for the males was determined to be 16. For the females K item selection ranged from a low of 10 to a high of 57 with the mean falling at 32.

Hoyt's analysis of variance technique was used in determining K scale reliability. For the males reliability estimates of .734 and .799 were obtained on validation and cross-validation samples, respectively. For the females reliability coefficients of .884 and .856 were determined respectively for validation and cross-validation samples.

To obtain evidence of K scale validity two approaches were examined for the effect of K on 1) test reliability and 2) test validity.

Individuals possessing K item T-scores of 65 or greater were excluded from the sample. Hoyt's analysis of variance technique for estimating internal consistency was used to obtain a reliability statement of the M-Scales before and after the application of the K scale. It was hypothesized that evidence of the effectiveness of the K scale could be determined by its ability to remove defensive individuals who tend to lower instrument reliability by erratic defensive test performance due to anxiety stemming from a threatening test situation. Theoretically, reliability should increase with exclusion of unreliable subjects. However, the effect of homogeneity of test performance may operate also to reduce reliability. The question was raised as to which has the greater effect on reliability: erratic, defensive test performance or homogeneity of test performance. To test the effects of the above question a random sample of subjects and equal in magnitude to those identified by high K scores were excluded. The assumption was made that the internal consistency reliability coefficient reduced by random selection should be greater than the reliability coefficient reduced by K selection.

For both male and female validation and cross-validation samples no significant differences in reliability coefficients were obtained, although in all cases the coefficient increased in magnitude after the application of K. With random exclusion all coefficients decreased in magnitude although none significantly.

The effects on validity correlation of M-Scales raw scores and standardized grade point averages before and after application of the male and female K scales were determined. Before application of K the male validation and cross-validation validity coefficients were .573 and .493, respectively. After use of the K scale the male correlations increased to .610 and .531 for the validation and cross-validation groups, respectively. For the females validation and cross-validation coefficients of .487 and .574 were obtained before use of K. After application of the

K scale these correlations increased to .556 and .611, respectively. However, no significant differences in correlations were obtained after K was applied. All validity correlations were significant from zero at the 3 percent or better level of confidence.

The concept of the K scale was corroborated, however, from the standpoint of the rationale of defensive test-taking attitudes. It was expected that the misclassified over-achiever who in essence is an under-achiever would be more threatened by the test situation and hence, would be more defensive. With such an attitudinal overlay this individual would choose more K items. This expectancy was significantly found to be the case for both males and females in the sample.

Conclusions

The following conclusions are based upon the findings of the investigation:

1. Females selected approximately two times as many defensive items as males. This suggests that the female from a societal point of view has more to be threatened by in the test situation than the male. Perhaps because of the greater range of interests and activities of the male in the society together with greater freedom of movement and lesser expectation to conform to a prescribed behavioral role, the male perceives less threat in test performance than the female.

2. Because of the sexual differences in patterns of K item selection, the K scale possesses the ability to tap an academic masculinity-femininity continuum.

3. The hypothesis that misclassified over-achievers would select more K items than misclassified under-achievers

was significantly substantiated. The misclassified over-achiever selected significantly more K items because of the anxiety overlay produced in him by a threatening test situation which consequently generates a defensive test-taking attitude.

4. The hypothesis that M-Scales reliability would increase significantly after exclusion of high K score individuals was not substantiated. As a consequence, the assumption that correlation reduced by random selection should be significantly greater than correlation reduced by K selection was not supported. Although such significant increases were not encountered there were still appreciable consistent increases in reliability coefficients when K was employed. In addition, when random exclusion was carried out an appreciable but not significant decrease in reliability was experienced. This suggests, but not strongly, that exclusion of defensive, erratic performance on a test is more influential on test reliability than the homogeneity concept of test performance.

5. The hypothesis that the correlation between M-Scales raw scores and standardized grade point averages would increase significantly after application of the K scale was not substantiated. Nevertheless, validity coefficients increased consistently for both males and females after use of K.

6. Further investigation with the K scale should be conducted before actual employment of the scale in test battery interpretation. Although there is significant empirical evidence to show that the scale can differentiate misclassified over-achievers from misclassified under-achievers, complete validation of the instrument is lacking.

7. The concept underlying the K scale of the MMPI is substantiated. Defensive test performance is a basis for a validation key. In addition, however, the K concept possesses the potential for a discriminating scale between various other subtle behavioral phenomena.

Implications for Further Research

Implications for further investigation include:

1. Development of a corrective K ratio score which when added to or subtracted from an individual's over all battery score would correct for defensive or over-truthful test performance.
2. Factorial analysis of the items comprising the male and female K scale.
3. Decrease the K item cut-off T-score from 65 to 60 and study the effects on test reliability and validity.
4. Examine behavioral characteristics of individuals selecting a high frequency of defensive items.
5. Construct and validate a research masculinity-femininity scale using male and female K scale items.

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

Edwards, A. L. Edwards Personal Preference Schedule. New York: The Psychological Corporation, 1957, 27 pp.

Fricke, B. G. The Opinion, Attitude and Interest Survey, Minneapolis: Investors Diversified Services, 1955.

Hartshorne, Hugh and May, M. A. Studies in Deceit. New York: Macmillan, 1928. 248 pp.

Hathaway, S. R. Supplementary Manual for the MMPI. New York: The Psychological Corporation, 1946.

Hathaway, S. R. and McKinley, J. C. The Minnesota Multiphasic Personality Schedule. Minneapolis: University of Minnesota Press, 1942.

Hathaway, S. R. and McKinley, J. C. Manual for the MMPI. New York: The Psychological Corporation, 1946.

Maller, J. B. Character Sketches. New York: Bureau of Publications, Teachers College, Columbia University, 1932. 388 pp.

Maller, J. B. "Personality Tests." in J. M. Hunt, Personality and the Behavior Disorders. New York: Ronald Press, 1944.

Kuder, G. F. Kuder Preference Record Vocational. Chicago: Science Research Associates, 1956. 35pp.

Kuder, G. F. Kuder Preference Record Occupational. Chicago: Science Research Associates, 1959. 18 pp.

Ruch, F. L. "A Technique for Detecting Attempts to Fake Performance on a Self-Inventory Type of Personality Test." In Quinn McNemar and M. A. Merrill, Studies in Personality. New York: McGraw-Hill, 1942.

Strong, E. K. Vocational Interest of Men and Women. Stanford: Stanford University Press, 1943. 746 pp.

Symonds, P. M. Diagnosing Personality and Conduct. New York: Appleton-Century, 1932. 602 pp.

B. PERIODICALS

Adams, C. R. "A New Measure of Personality," Journal of Applied Psychology, 1941, 25:141-151.

Allport, G. W. "A Test for Ascendance-Submission," Journal of Abnormal Psychology, 1928, 23:118-136.

Allport, G. W. "The Use of Personal Documents in Psychological Science," Social Science Research Council Bulletin, 1942, Number 42.

Benton, A. L. "The MMPI in Clinical Practice," Journal of Nervous and Mental Disorders, 1945, 102:416-420.

Benton, A. L. "The Interpretation of Questionnaire Items in a Personality Inventory," Archives of Psychology, 1935, Number 190.

Bernreuter, R. G. "Validity of the Personality Inventory," Personality Journal, 1933, 11:383-386.

Bernreuter, R. G. "The Present Status of Personality Trait Tests," Educational Research Supplement, 1940, 21:160-171.

Bernreuter, R. G. "Theory and Construction of the Personality Inventory," Journal of Social Psychology, 1933, 4:387-405.

Bills, Marion. "Selection of Casualty and Life Insurance Agents," Journal of Applied Psychology, 1941, 25:6-10.

Bordin, E. S. "A Theory of Vocational Interests as Dynamic Phenomena," Educational and Psychological Measurement, 1943, 3:49-65.

Cady, V. M. "The Estimation of Juvenile Incorrigibility," Journal of Delinquency Monographs, 1923, Number 2.

Cofer, C. N., Chance, J. and Judson, A. J. "A Study of Malingering on the MMPI," Journal of Psychology, 1949, 27:491-499.

Cottle, W. C. "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, 1950, 34:255-259.

Cottle, W. C. "The MMPI: A Review," Kansas Studies in Education, 1953, 3:1-82.

Cronbach, L. J. "Response Sets and Test Validity," *Educational and Psychological Measurements*, 1946, 6:475-494.

Cronbach, L. J. "Further Evidence on Response Sets and Tests Designs," *Educational and Psychological Measurement*, 1950, 10:3-31.

Eisenberg, P. "Individual Interpretation of Psychoneurotic Inventory Items," *Journal of Genetic Psychology*, 1941, 25:19-40.

Eisenberg, P. and Wesman, A. "A Consistency in Responses and Logical Interpretation of Psychoneurotic Inventory Items," *Journal of Educational Psychology*, 1941, 32:321-338.

Frenkel-Brunswik, E. "Mechanisms of Self-Deception," *Journal of Social Psychology*, 1939, 10:409-420.

Fricke, B. G. "A Response Bias Scale for the MMPI," *Journal of Counseling Psychology*, 1957, 4:149-153.

Fricke, B. G. "Conversion Hysterics and the MMPI," *Journal of Clinical Psychology*, 1956, 12:322-326.

Fricke, B. G. "Subtle and Obvious Test Items and Response Set," *Journal of Consulting Psychology*, 1957, 21:250-252.

Gough, H. G. "Simulated Patterns on the MMPI," *Journal of Abnormal and Social Psychology*, 1947, 42:215.

Gough, H. G. "Factors Relating to the Academic Achievement of High School Students," *Journal of Educational Psychology*, 1949, 40:65-78.

Gough, H. G. "What Determines the Academic Achievement of High School Students," *Journal of Educational Research*, 1953, 46:321-331.

Gough, H. G. "The F Minus K Dissimulation Index for the MMPI," *Journal of Consulting Psychology*, 1950, 14:408-413.

Guilford, J. P. and Guilford, R. B. "Personality Factors S, E, and M and Their Measurement," *Journal of Psychology*, 1936, 2:109-127.

Hathaway, S. R. and McKinley, J. C. "A Multiphasic Personality Schedule: I. Construction of the Schedule," *Journal of Psychology*, 1940, 10:249-254.

Hathaway, S. R. and McKinley, J. C. "A Multiphasic Personality Schedule: III. The Measurement of Symptomatic Depression," Journal of Psychology, 1942, 14:73-84.

Horst, Paul. "The Prediction of Personal Adjustment," Social Science Research Council Bulletin, 1941, Number 48.

Hovey, H. B. "Detection of Circumvention in the MMPI," Journal of Clinical Psychology, 1948, 4:97.

Hoyt, C. J. "Test Reliability Estimated by Analysis of Variance," Psychometrika, 1941, 6:133-160.

Humm, D. G. and Humm, K. A. "Validity of the Humm-Wadsworth Temperament Scale: With Consideration of the Effects of Subjects' Response-Bias," Journal of Psychology, 1944, 18:55-64.

Humm, D. G., Storment, R. C. and Iorns, M. E. "Combination Scores for the Humm-Wadsworth Temperament Scale," Journal of Psychology, 1939, 7:227-253.

Humm, D. G. and Wadsworth, G. W. "The Humm-Wadsworth Temperament Scale," American Journal of Psychiatry, 1935, 92:163-200.

Hunt, H. F. "A Study of the Differential Diagnostic Efficiency of the MMPI," Journal of Consulting Psychology, 1948, 12:331-336.

Hunt, H. F. "The Effect of Deliberate Deception on the MMPI Performance," Journal of Consulting Psychology, 1948, 12: 396-402.

Hunt, W. A. "The Detection of Malingering: A Further Study," United States Naval Medical Bulletin, 1946, 46-249.

Hunt, W. A. and Older, H. J. "Detection of Malingering Through Psychometric Tests," United States Naval Medical Bulletin, 1943, 41:1318.

Kazan, A. T. and Sheinberg, I. M. "Clinical Note on the Significance of the Validity Score F in the MMPI," American Journal of Psychiatry, 1945, 102:181-183.

Kelly, E. L., Miles, C. C. and Terman, L. M. "Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality," Character and Personality, 1936, 4:206-215.

Krumboltz, J. D. and Farquhar, W. W. "The Effect of Three Teaching Methods on Achievement and Motivational Outcomes in a How-to-Study Course," Psychological Monographs, 1957, 71: Number 14.

Laird, D. A. "Detecting Abnormal Behavior," Journal of Abnormal Psychology, 1926, 20:128-141.

Landis, C. and Katz, S. E. "The Validity of Certain Questions With Purport to Measure Neurotic Tendencies," Journal of Applied Psychology, 1934, 18:343-356.

Levine, A. S. "A Technique for Developing Suppression Tests," Educational and Psychological Measurement, 1952, 12:313-315.

Maller, J. B. "The Effect of Signing One's Name," School and Society, 1930, 31:882-884.

McKinley, J. C. and Hathaway, S. R. "A Multiphasic Personality Schedule: V. Hysteria, Hypomania and Psychopathic Deviate," Journal of Applied Psychology, 1942, 14:73-84.

McKinley, J. C., Hathaway, S. R. and Meehl, P. E. "The MMPI: VI. The K Scale," Journal of Consulting Psychology, 1948, 12:20-31.

McNemar, Quinn. "The Mode of Operation of Suppressant Variables," American Journal of Psychology, 1945, 58:554-555.

McQuary, J. J. and Traux, W. E. "An Under Achievement Scale," Journal of Educational Research, 1955, 48:393-399.

Meehl, P. E. "An Investigation of General Normality Control Factor in Personality Testing," Psychological Monographs, 1945, 59:Number 4.

Meehl, P. E. "The Dynamics of Structured Personality Tests," Journal of Clinical Psychology, 1945, 1:296-303.

Meehl, P. E. and Hathaway, S. R. "The K Factor as a Suppressor Variable in the MMPI," Journal of Applied Psychology, 1946, 30:525-564.

Metfessel, M. "Personality Factors in Motion Picture Writing," Journal of Social and Abnormal Psychology, 1935, 30:333-347.

Middleton, George and Gutherie, G. M. "Personality Syndromes and Academic Achievement," Journal of Educational Psychology, 1959, 50:Number 2.

Mosier, C. I. "A Note on Item Analysis and the Criterion of Internal Consistency," Psychometrika, 1936, 1:275-282.

Olson, W. C. "The Waiver of Signature in Personal Reports," Journal of Applied Psychology, 1936, 20:442-450.

Ossipov, V. P. "Malingering: The Simulation of Psychosis," Bulletin of the Menninger Clinic, 1944, 8:33-42.

Rosen, E. "Self Appraisal and Perceived Desirability of MMPI Personality Traits," Journal of Counseling Psychology, 1956, 3:44-51.

Rosen, E. "Self Appraisal, Personal Desirability, and Perceived Social Desirability of Personality Traits," Journal of Abnormal and Social Psychology, 1956, 52:151-158.

Rosenzweig, Saul. "A Suggestion for Making Verbal Personality Tests More Valid," Psychological Review, 1934, 41:400-401.

Rosenzweig, Saul. "A Basis for the Improvement of Personality Tests with Special Reference to the M-F Battery," Journal of Abnormal and Social Psychology, 1938, 33:476-488.

Schmidt, H. O. "Test Profiles as a Diagnostic Aid: The MMPI," Journal of Applied Psychology, 1945, 29:115-131.

Schmidt, H. O. "Notes on the MMPI: The K Factor," Journal of Consulting Psychology, 1948, 12:337-342.

Schneck, J. M. "Clinical Evaluation of the F Scale on the MMPI," American Journal of Psychiatry, 1948, 104:440-442.

Shoben, E. J. "The Assessment of Parental Attitudes in Relation to Child Adjustment," Genetic Psychology Monographs, 1949, 39:101-148.

Spencer, D. "Frankness of Subjects on Personality Measures," Journal of Educational Psychology, 1938, 29:26-35.

Steinmetz, H. C. "Measuring Ability to Fake Occupational Interest," Journal of Applied Psychology, 1932, 16:123-230.

Sweetland, A. "Hypnotic Neurosis-Hypochondriasis and Depression," *Journal of Genetic Psychology*, 1948, 39:19-105.

Vernon, P. E. "The Attitude of the Subject in Personality Testing," *Journal of Applied Psychology*, 1934, 18:165-177.

Washburne, J. N. "A Test of Social Adjustment," *Journal of Applied Psychology*, 1935, 19:125-244.

Wherry, R. J. "Test Selection and Suppressor Variables," *Psychometrika*, 1946, 11:239-247.

Wiener, D. N. "Selecting Salesmen with Subtle-Obvious Keys for the MMPI," *American Psychologist*, 1948, 3:364.

Wiener, D. N. "Subtle and Obvious Keys for the MMPI," *Journal of Consulting Psychology*, 1948, 12:164-170.

Wiener, D. N. "A Control Factor in Social Adjustment," *Journal* of Abnormal and Social Psychology, 1951, 46:3-8.

Willoughby, R. R. and Morse, M. E. "Spontaneous Reactions to a Personality Inventory," *American Journal of Orthopsychiatry*, 1936, 6:562-575.

C. UNPUBLISHED MATERIALS

Arnold, D. A. "The Clinical Validity of the Humm-Wadsworth Temperament Scale in Psychiatric Diagnosis," Unpublished Doctor's dissertation, University of Minnesota, Minneapolis, 1942.

Farquhar, William W. "Motivation Factors Related to Academic Achievement," East Lansing: Cooperative Research Project No. 846, Office of Education, United States Department of Health, Education and Welfare, January, 1963.

Fricke, B. G. "The Development of an Empirically Validated Personality Test Employing Configural Analysis for the Prediction of Academic Achievement," Unpublished Doctor's dissertation, University of Minnesota, 1954.

Hendrickson, G. "Attitudes and Interests of Teachers and Prospective Teachers." Paper read before Section Q, AAAS, Atlantic City, December, 1932.

Jeffery, M. E. "Some Factors Influencing Answers on the Multiphasic K Scale." Unpublished Doctor's dissertation, University of Minnesota, Minneapolis, 1946.

APPENDIX

APPENDIX A

K Scale Female Items, Significance Level and Scoring Direction

Item Number

Generalized Situational Choice Inventory	Statistical Significance .05 Level or Higher	Direction of Scoring Answer Number
4	6.067	1
32	4.809	1
41	6.510	1
42	8.041	2
54	4.809	1
56	5.947	1
58	4.438	1
96	3.967	2
101	6.856	2
107	[illegible]	1
112	7.769	1
125	5.463	1
132	5.453	2
141	10.599	2
164	11.737	1
174	8.366	2

Word Rating List

7	6.092	2
8	5.928	2
10	6.074	2
13	4.835	2
15	5.339	2
16	5.922	2
18	13.841	2
22	5.400	2
25	6.081	2
29	7.459	2
30	9.874	2
33	11.991	2
36	4.835	2
46	13.841	2
47	11.785	2
52	5.612	2
56	8.024	2

APPENDIX A

K Scale Female Items, Significance Level and Scoring Direction

Item Number

Generalized Situational Choice Inventory	Statistical Significance .05 Level or Higher	Direction of Scoring: Answer Number
4	6.067	1
33	4.809	1
41	6.510	1
42	8.041	2
54	4.809	1
56	5.947	1
58	4.458	1
96	3.967	2
101	6.854	2
107	6.782	1
112	7.769	1
125	8.463	1
132	5.653	2
141	10.559	2
164	11.757	1
174	6.346	2
Word Rating List		
7	6.092	2
8	5.928	2
10	4.074	2
13	4.835	2
15	5.339	2
16	3.922	2
18	13.841	2
22	5.400	2
25	6.081	2
29	7.459	2
30	9.874	2
33	11.991	2
34	4.835	2
46	13.841	2
47	11.765	2
52	5.612	2
56	8.074	2

APPENDIX A (continued)

Word Rating List

57	9.680	2
67	4.259	2
68	4.883	2
71	9.397	2
75	5.928	2
77	4.214	2
78	9.442	2
80	4.918	2
83	12.055	2
88	6.486	2
97	6.084	2
98	6.253	2
105	5.931	2
108	11.620	2
118	31.650	2

Human Trait Inventory

17	5.154	2
22	3.847	2
26	3.852	1 or 2
27	11.360	2
42	6.012	1 or 2
43	10.694	2
60	4.056	1 or 2
68	3.847	3 or 4
77	10.164	3 or 4
100	3.847	1 or 2
110	6.235	1 or 2
117	6.470	1 or 2
122	7.496	2

Preferred Job Characteristics Scale

5	4.928	1
11	9.365	2
12	10.264	2
13	3.847	1
18	8.318	2
19	12.724	1
21	5.929	1

APPENDIX A (continued)

Preferred Job Characteristics Scale

25	4.802	2
26	8.474	2
27	8.413	1
33	9.365	1
34	4.863	1
35	6.783	1
37	8.413	1
38	21.130	2
42	4.250	2
43	7.132	2
44	4.863	1
45	9.868	1
51	8.041	1
52	5.199	1
55	6.783	1
59	12.111	1
60	3.847	1
61	8.413	2
62	10.697	1

APPENDIX B

K Scale Male Items, Significance Level, and Scoring Direction

Item Number

Generalized Situational Choice Inventory	Statistical Significance .05 Level or Higher	Direction of Scoring Answer Number
26	5.756	2
37	7.057	1
61	5.141	1
77	4.572	1
91	5.397	1
143	4.952	1
145	3.964	2
192	7.057	1
Word Rating List		
12	5.207	1 or 2
13	4.987	2
14	8.755	3 or 5
15	8.755	2
16	13.934	2
17	5.827	2
18	4.644	2
30	4.007	1 or 2
34	4.952	1 or 2
36	8.349	2
38	4.225	2
41	7.170	2
67	11.267	2
68	4.759	2
83	4.644	2
96	11.778	1 or 2
98	5.141	2
109	4.693	2
117	4.644	3
119	3.987	2

APPENDIX B (continued)

Human Trait Inventory

47	3.972	1 or 2
48	4.212	1 or 2
73	6.603	3 or 4
107	4.674	1 or 2
108	10.961	1 or 2
116	4.764	1 or 2
117	9.357	1 or 2

Preferred Job Characteristics Scale

5	6.506	1
9	4.128	2
19	6.548	1
33	7.549	1
35	4.070	1
45	5.397	1
52	4.766	1